Biochemical Characterization of the Phoslactomycin modular Polyketide Synthase

Dissertation

kumulativ

zur Erlangung des Grades eines

Doktor der Naturwissenschaften

(Dr. rer. nat.)

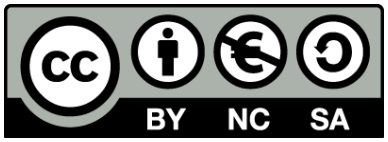
des Fachbereichs Biologie der Philipps-Universität Marburg

vorgelegt von

Kyra Geyer

aus Brilon

Marburg, 2020



Die Arbeit zur vorliegenden Dissertation wurden von Februar 2017 bis Oktober 2020 unter der Betreuung von Herrn Prof. Dr. Tobias J. Erb in Marburg am Max-Planck-Institut für terrestrische Mikrobiologie in der Abteilung Biochemie und Synthetischer Metabolismus durchgeführt.

Vom Fachbereich Biologie der Philipps-Universität Marburg (Hochschulkennziffer 1180) als Dissertation angenommen am 16.12.2020

Erstgutachter: Prof. Dr. Tobias J. Erb
Zweitgutachter: Prof. Dr. Lars-Oliver Essen

Tag der Disputation: 15.01.2021

Erklärung

Ich versichere, dass ich meine Dissertation mit dem Titel „Biochemical Characterization of the Phoslactomycin modular Polyketide Synthase“ selbstständig ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen und Hilfsmittel bedient habe.

Diese Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Marburg, den

Kyra Geyer

Alles Wissen und alles Vermehren unseres Wissens
endet nicht mit einem Schlußpunkt,
sondern mit einem Fragezeichen.

Hermann Hesse (1877-1962), dt. Dichter, 1946 Nobelpr. f. Lit.

Parts of this thesis have been published or are in preparation for publication:

Bastian Vögeli, **Kyra Geyer**, Patrick D. Gerlinger, Sarah Benkstein, Niña Socorro Cortina, Tobias J. Erb. Combining Promiscuous Acyl-CoA Oxidase and Enoyl-CoA Carboxylase/Reductases for Atypical Polyketide Extender Unit Biosynthesis. ***Cell Chemical Biology 2018,*** *25, 1-7*

In this work K.G. contributed to: The design and performance of experiments, the analysis of data, the writing of the manuscript

Kyra Geyer, Srividhya Sundaram, Peter Sušnik, Ulrich Koert, Tobias J. Erb. Understanding Substrate Selectivity of Phoslactomycin Polyketide Synthase by Using Reconstituted in Vitro Systems. ***ChemBioChem 2020****, 21, 14*

In this work K.G. contributed to: the conception of the project, the design and performance of experiments, the design and establishment of the steady state kinetic assay, the design and the establishment of the *in vitro* PKS system, the analysis of data, the writing of the manuscript

Kyra Geyer, Steffen Hartmann, Randolph R. Singh, Tobias J. Erb. Function of the type II thioesterase associated with the phoslactomycin polyketide synthase. *(In preparation)*

In this work K.G. contributed to: the conception of the project, the design and performance of experiments, the analysis of data, the writing of the manuscript

The above stated contributions of K.G. to the publications or manuscripts in preparation, which are part of this thesis, are herewith endorsed by:

Kyra Geyer Prof. Dr. Tobias Erb

Publication not discussed in this thesis:

Alexandra A. Richter, Christopher-Nils Mais, Laura Czech, **Kyra Geyer**, Astrid Hoeppner, Sander H.J. Smits, Tobias J. Erb, Gert Bange and Erhard Bremer (2019). Biosynthesis of the Stress-Protectant and Chemical Chaperon Ectoine: Biochemistry of the Transaminase EctB. ***Front. Microbiol****.* 10:2811.

Table of Contents

List of abbreviations

ACP	Acyl carrier protein
AT	Acyltransferase
CoA	Coenzyme A
Cryo-EM	Cryogenic electron microscopy
DEBS	Erythromycin PKS, generates 6 deoxyerythronolide B
DH	Dehydratase
ECR	Enoyl-CoA carboxylase/reductase
FACL	Fatty acyl-CoA ligases
FAS	Fatty acid synthase
KR	Ketoreductase
KS	Ketosynthase
NRPS	Non ribosomal peptide synthetase
PKS	Polyketide synthase
Pn PKS	Phoslactomycin polyketide synthase
PnA_{V4}	First enzyme of Pn PKS, N-terminal truncated
PnB, $PnB\text{-}TE_{DEBS}$	Second enzyme of Pn PKS without and with terminal TE
PnC, $PnC\text{-}TE_{DEBS}$	Third enzyme of Pn PKS without and with terminal TE
PnD, $PnD\text{-}TE_{DEBS}$	Fourth enzyme of Pn PKS without and with terminal TE
PnG	Thioesterase from the phoslactomycin PKS
TE, TEI, TEII	Thioesterase, thioesterase type I, thioesterase type II
TE_{DEBS}	Terminal thioesterase from DEBS PKS

Summary

Polyketides are a class of natural products with a large structural diversity. They find use for example in infection-, cancer- and respiratory disease treatments. New discoveries, chemical modification and the engineering of polyketide biosynthetic pathways may lead to the identification of novel products with altered, potentially improved properties.

The structural diversity of polyketides originates from the assembly line-like joining of simple building blocks by polyketide synthases (PKS). The introduction of structural features is guided by the choice of starter units and the incorporation of various extender units, such as malonyl- or alkyl-malonyl-CoAs. Enoyl-CoA carboxylases/reductases (ECRs) are the key-enzymes providing alkyl-malonyl-CoAs. ECR catalyzed carboxylation of *α,β*-unsaturated enoyl-CoA thioesters, yields in the production of e.g. (2*S*)-ethyl-malonyl-CoA. This work presents a route for the preparative scale chemo-enzymatic synthesis of a versatile set of extender units. Combining ECR activity with the concept of biocatalytic proof-reading for recycling of unwanted byproducts resulted in the efficient production of various polyketide extender units.

This set of extender units enabled tackling the fundamental question, of how site-specific incorporation of different extender units in PKS is conveyed. For that, phoslactomycin PKS (Pn PKS) derived tetra-, penta- and hexaketide were produced *in vitro*. Furthermore, challenging the Pn PKS assembly line, that naturally incorporates malonyl- and ethylmalonyl-CoA, with seven different extender units, revealed a highly promiscuous module. With detailed kinetic analysis of excised domains, it could be shown that the transacylation reaction is the driving force determining the incorporation of extender units.

Furthermore, in this work, the Pn PKS associated type II thioesterase PnG could be shown to possess an editing function and thereby increase the yield of native and non-native phoslactomycin derived polyketides *in vitro*. This work underlines their potential for improving the production yields of polyketides.

In summary, a detailed study of the phoslactomycin PKS, starting from the production of extender units, over the *in vitro* reconstitution of Pn PKS, to the characterization of single domains and an accessory thioesterase, is presented. This dissertation provides new insights in the molecular functions of this astonishing class of enzymes.

Zusammenfassung

Polyketide sind Sekundärmetabolite mit einer großen strukturellen Vielfalt. Sie werden beispielsweise als Mittel gegen Infektionen, Krebs oder Atemwegserkrankungen eingesetzt. Neuentdeckungen, chemische Modifikationen und Engineering von Polyketidbiosynthesewegen kann zu neuen Produkten mit veränderten, potentiell besseren Eigenschaften führen.

Die strukturelle Vielfalt von Polyketiden entsteht durch die schrittweise Kopplung von einfachen Bausteinen durch Polyketid-Synthasen (PKS), wobei die Auswahl der Bausteine, wie Malonyl- und Ethylmalonyl-CoA zum Einbau von strukturellen Merkmalen führt. Die Schlüsselenzyme zur Darstellung von Alkylmalonyl-CoAs sind die Enoyl-CoA Carboxylasen/Reduktasen (ECRs). Diese Enzyme katalysieren die Carboxylierung von *α*,*β*-ungesättigten Enoyl-CoA Estern, was in der Produktion von z.B. (2*S*)-Ethylmalonyl-CoA resultiert. In dieser Arbeit wird eine preparative chemo-enzymatische Synthese für diverse Bausteine präsentiert. Die ECR Aktivität wurde mit dem Konzept des biokatalytischen Korrekturlesens zum Wiederverwerten ungewollter Nebenprodukte kombiniert, was die effiziente Produktion einer Auswahl von Polyketid-Bausteinen ermöglicht.

Mit den so hergestellten Bausteinen, konnte der grundlegenden Frage, wie der Einbau verschiedener Bausteine vermittelt wird, nachgegangen werden. In dieser Arbeit wurden Tetra-, Penta- und Hexaketide des Phoslactomycin PKS (Pn PKS) *in vitro* produziert. Das Pn PKS System wurde mit sieben verschiedenen Bausteinen konfrontiert, was zur Identifikation eines promiskuitiven Moduls führte. Detaillierte kinetische Untersuchung einzelner Domänen zeigte, dass die Transacylierungsreaktion der entscheidende Faktor ist, der zum Einbau von Bausteinen führt.

Zusätzlich wurde in dieser Arbeit gezeigt, dass die Pn PKS assoziierte Typ II Thioesterase PnG, eine PKS-reinigende Funktion aufweist, die zur höheren Produktion von nativen und nicht nativen Pn Polyketiden *in vitro* führt. Dies unterstreicht das Potential dieser Enzyme für die Produktion von biosynthetisch veränderten Polyketiden.

Zusammenfassend ist diese Dissertation eine detaillierte Untersuchung der Phoslactomycin PKS, beginnend bei der Produktion versatiler Bausteine, über die *in vitro* Rekonstruierung der Pn PKS, bis hin zur Charakterisierung der zugehörigen Thioesterase und präsentiert neue Einblicke in diese beeindruckende Enzymklasse.

Introduction

History of natural products

Natural products have been used by humans in traditional medicine for centuries. Some of the first written records found in Mesopotamia are dating back to the year 2600 BC and are collections of information on plant derived substances. Some of these compounds are still used in modern medicine for treatments of coughs, parasitic infections and inflammations [1]. These extracts, often presented as oils were mixtures of multiple compounds. The effort to identify single active compounds started in the early 19th century with the purification of morphine, atropine and colchicine. Soon after the first commercialization of pure natural products, semi-synthetic drugs like aspirin became available [1].

The applications of natural products in human medicine cover many roles including fighting infections, lowering cholesterol levels and treating cancer [2, 3]. Besides in human medicine, they are also applied in veterinary medicine and agriculture [4]. In the past decades the number of natural products or natural product-derived drugs has risen to roughly 34% of all drugs on the market with an additional 20% of synthetic drugs containing natural product pharmacophores. In the field of antibacterials this number even reaches 55% [5], underlining their indispensability in modern medicine.

Natural products are omnipresent, as they are found all over the tree of life in bacteria, fungi, plant and animals [4, 6]. In the kingdom of bacteria, the family Actinomycetes has proven to be a rich source of various compounds. Two thirds of the known antibiotics are produced by them, the majority by *Streptomycetes* [7]. Depending on the biosynthetic origin, natural products are divided into polyketides, non-ribosomal peptides, terpenes, alkaloids and lanthipeptides amongst others. All these compounds are produced by secondary metabolism and are thus not required for survival itself, nevertheless they provide advantages for the producer in the native environment. The biological function of these compounds is diverse but not limited to defense - serving as toxins and antibiotics, communication, quorum sensing, virulence and pigmentation [2, 8, 9].

Polyketides comprising of polyethers, polyenes, polyphenols, macrolides and enediynes are one of the major classes of natural products and a significant source of new drugs and bioactive small molecules. Some representatives such as erythromycin,

doxorubicin, epothilone B, ivermectin and lovastatin, which are used as antibiotic, antitumor, anti-parasitic and cholesterol lowering agents are blockbuster drugs with a great impact on modern civilization [10].

Biosynthesis of polyketides

Polyketide synthases (PKS) and fatty acid synthases (FAS) share a common evolutionary history and thus show similar biosynthetic characteristics in precursor usage, domain organization and catalytic mechanism [3, 11, 12]. The products formed by both enzyme classes are assembled in an assembly line fashion by the repetitive action of catalytic units (termed domains). The acyltransferase (AT) selects the substrate and transfers it from CoA to the thiol group of the phosphopantetheine arm of the cognate acyl-carrier protein (ACP). This prosthetic group serves as a flexible arm that shuttles the growing polyketide chain between domains. The *β*-ketoacylsynthase (KS) catalyzes a C-C bond formation between the acyl-unit bound to a cysteine thiol and the extender unit bound to the ACP. This condensation results in a two-carbon head-to-tail elongation of the growing polyketide. By the irreversible elimination of CO_2 a reactive enolate anion is formed, which serves as a nucleophile to extend the polyketide. This reaction is termed decarboxylative Claisen condensation. Following chain elongation, the *β*-keto group can be successively reduced by a ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER), finally resulting in a saturated acyl-chain. In some cases, a methyltransferase (MT) is embedded in the assembly line and can add an *α*-methyl branch [2, 13]. Multiple repetitions of this cycle are performed until the full length of the product is reached, which is then typically released by a thioesterase (TE) (Figure 1).

The evolutionary separation of FAS and PKS represents a branching point between primary and secondary metabolism [2]. Besides the fundamental similarities between these two protein classes, there are also characteristic differences that lead to the immense structural diversity of polyketides compared to fatty acids. PKS use a broader range of substrates, going beyond acetyl- and malonyl-CoA commonly used by FAS. Furthermore, they use different mechanisms to control product chain length and in contrast to FAS, the reductive steps are optional giving rise to multiple functional groups.

Assembly of the polyketide backbone thus sets the basis of structural diversity by variable chain lengths, incorporation of *α*-substituted extender units, extent of *β*-reduction (resulting in *β*-keto, *β*-hydroxy, *α,β*-olefinic or *β*-CH_2 methylene groups) and installation of varying stereochemistry of the *α*-side chains, *β*-groups and double bonds resulting in an apparently infinite number of combinations. After biosynthesis of the carbon backbone, polyketides undergo extensive decoration such as oxidation, glycosylation, amination and alkylation, catalyzed by tailoring enzymes, adding further functionalities [3].

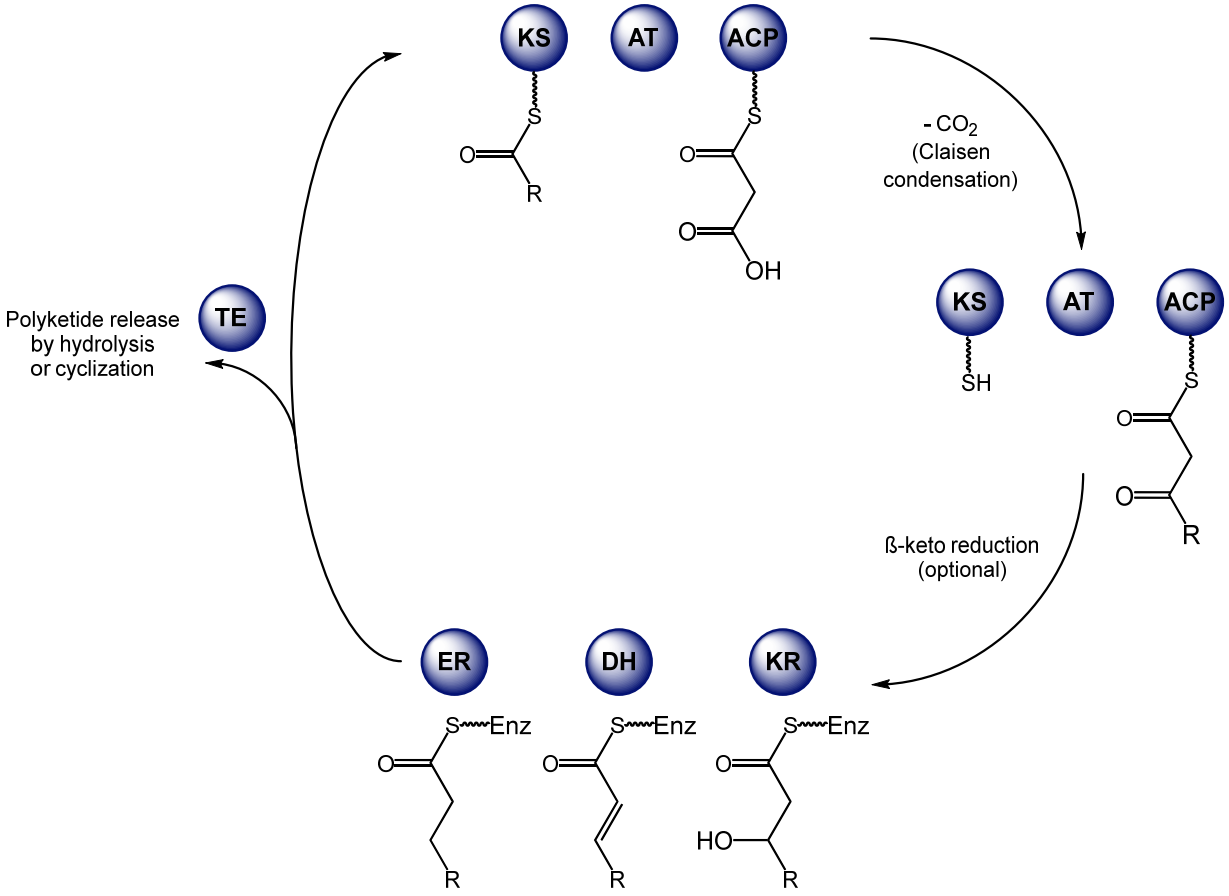


Figure 1: Basic steps in polyketide assembly with the involved domains. Ketoacylsynthase (KS), acyltransferase (AT), acyl carrier protein (ACP), ketoreductase (KR), dehydratase (DH), enoylreductase (ER) and thioesterase (TE).

Polyketide biosynthetic enzymes are divided into three classes. Type I PKS are exceptionally large, multifunctional, multi-domain enzymes with covalently linked domains

that are organized in modules (Figure 2). They are categorized into iterative and non-iterative PKS. In iterative modular PKS each module conducts multiple elongation and reduction reactions. In non-iterative modular PKS each module is responsible for a single elongation and a set of optional reduction steps. The assembly of their products follow the co-linearity rule, where the polyketide backbone can be predicted upon domain specificities and modules present [2, 3, 14-16].

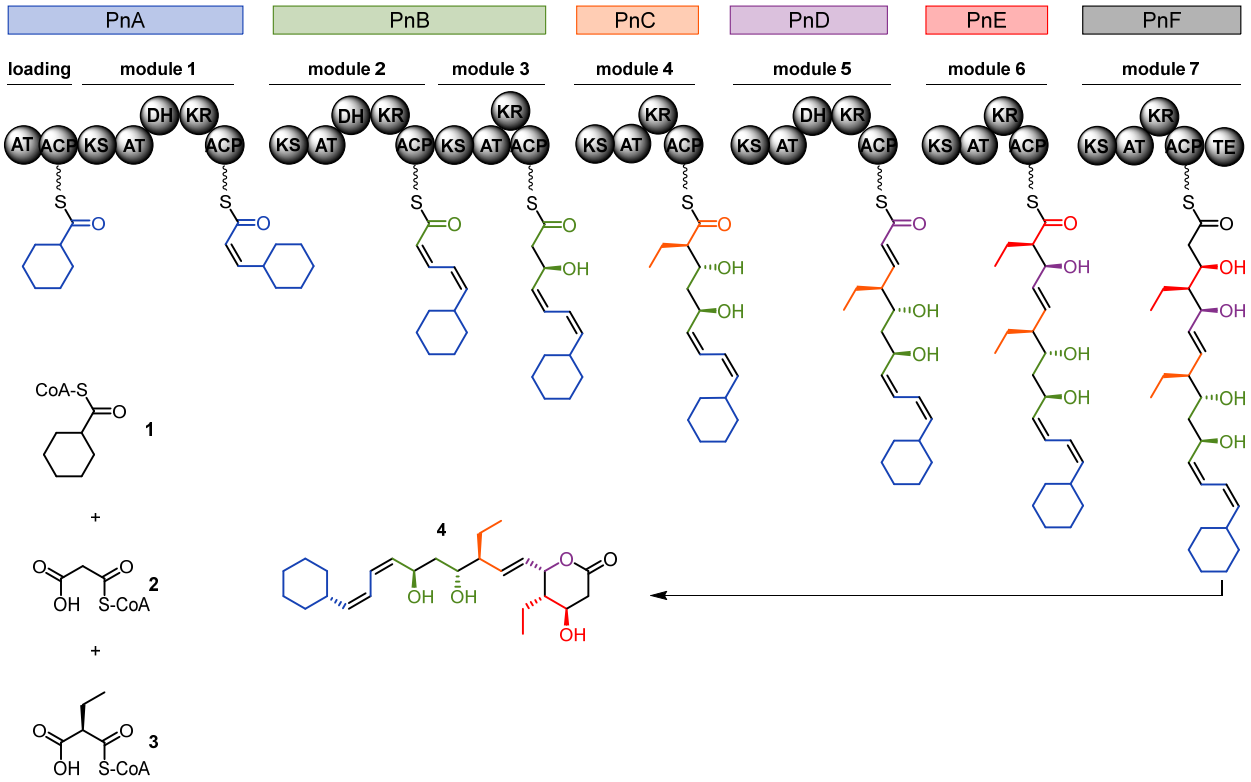


Figure 2: Phoslactomycin PKS from *Streptomyces platensis* as an example for a modular type I PKS system. The domains are organized in modules, that assemble to form the megasynthase complex. AT=Acyltransferase, ACP=Acyl-carrier protein, KS=Ketosynthase, DH=Dehydratase, KR=Ketoreductase. **1**=cyclohexanecarboxyl-CoA, **2**=malonyl-CoA, **3**=(2*S*)-ethylmalonyl-CoA. The polyketide is handed on from the first module PnA downstream to the last module PnF. After reaching full length the the octaketide (**4**) is released by lactonization catalyzed by the terminal TE (Thioesterase).

A special type of modular PKS are the *trans*-AT PKS, where one or multiple free-standing *trans*-acting ATs load the extender unit onto their ACP partner. This type of PKS generally does not follow the *cis*-AT PKS co-linearity rule. However, significant progress based on domain sequences has been made to predict the potential *trans*-AT

polyketide structure [17]. Type II PKS are discrete mono-functional enzymes that assemble into dissociable complexes and iteratively elongate the polyketide [18]. Type III PKS, also known as chalcone synthase type proteins, are multifunctional enzymes selecting the starter unit, assembling the product and cyclizing it, in an ACP-independent manner [19].

The structure of polyketide synthases

Until now, no crystal structure of a complete PKS module could be produced. Nevertheless, high resolution structures of multiple excised domains and some di-domains have been successfully solved. Remarkably, a cryogenic electron microscopy (cryo-EM) structure of the full pikromycin module (PikAIII) has been reported [20-26]. While this presents a great achievement towards final structure elucidation of modular PKS, the cryo-EM structure is debated and a precise picture of PKS module structure cannot been drawn, due to inconsistencies between the cryo-EM and crystal structures [27, 28].

A common feature of recent models is that the condensing domains (KS-AT) and the modifying domains (DH-ER-KR-ACP) are separated into two distinct regions as observed in metazoan FAS [29]. PKS are homodimeric enzymes with the KS, DH and TE domains being responsible for the dimerization. Additional short sequence motifs responsible for coiled-coil interactions, located at the *N*-terminus of the KS, KR and the *C*-terminus of the ACP have been identified to stabilize the homodimerization of PKS modules and the interaction with the downstream module [30, 31]. ER domains are members of the MDR superfamily, which predominantly exist as dimers or tetramers. Yet, data on the oligomeric state of PKS ER domains support the dimeric as well as a monomeric state [22, 23]. Between non-iterative and iterative modular PKS, the oligomeric state of the ER domain could be an important structural difference. In the crystal structure of the iterative mycocerosic acid synthase (MAS)-like PKS, which is closely related to modular PKS, the ER is found to be dimeric [32]. This structural feature is supported by the longer ER-KR linker sequences of iterative PKS compared to the linker sequences in modular PKS, which enables dimerization [27].

It was found that the DH dimer behaves as a rigid body, while the following ER domain exhibits slight flexibility, followed by the KR as the catalytic active domain with the highest flexibility [32]. During biosynthesis, the substrate is tethered to the ACP and shuffled

between the various domains. With the cryo-EM analysis of PikAIII a striking ACP movement depending on the phosphopantetheine-arm linked acyl-residue was observed [33]. A high-resolution structure of a PKS is necessary to fully understand the mechanism behind polyketide biosynthesis, domain interaction and ACP movement. This will result in the efficient guiding of the engineering of these highly complex systems.

Origin of the structural backbone diversity

An important mechanism accounting for the structural diversity of polyketides is the use of starter units other than acetyl-CoA, such as cyclohexanecarboxyl-CoA in phoslactomycin, p-aminobenzoic acid in candicidin, or isovaleryl-CoA in avermectin [34-36].

Following starter unit selection, most PKS use malonyl-CoA as an extender unit. In rarer cases, *α*-substituted extender units are used, leading to the incorporation of alkyl-residues at the at the *α*-carbon of polyketides. This is seen for example in the antibiotic erythromycin, which originates from the exclusive incorporation of methylmalonyl-CoA provided by the propionyl-CoA carboxylase [37]. With the growing number of polyketides characterized, more atypical *α*-substituted extender units were identified, such as ethylmalonyl-, benzylmalonyl-, hexylmalonyl- or allylmalonyl-CoA [38-41].

Most of these rare extender units are provided by enoyl-CoA carboxylase/reductases (ECRs). These enzymes are members of the medium-chain reductase/dehydrogenase (MDR) protein superfamily and catalyze the formation of a C-C bond between an *α,β*-unsaturated enoyl-CoA thioester and the electrophile CO_2. The reductive carboxylation is done in a stereospecific manner, resulting in (2*S*)-alkyl-malonyl-CoA derivatives suited for polyketide biosynthesis [42, 43]. ECRs are distinguished in two subfamilies, ECR-I from primary metabolism and ECR-II from secondary metabolism. While both can produce ethylmalonyl-CoA, PKS mainly draw their extender units from ECR-II activity [44]. These carboxylases show the side reaction of simply reducing the unsaturated CoA thioester without carboxylation, as a remnant of their evolutionary origin, resulting in saturated alkyl-chains (Figure 3). This is vividly demonstrated by the ECR AntE, which only carboxylates 25% of its natural substrate cinnamoyl-CoA [45].

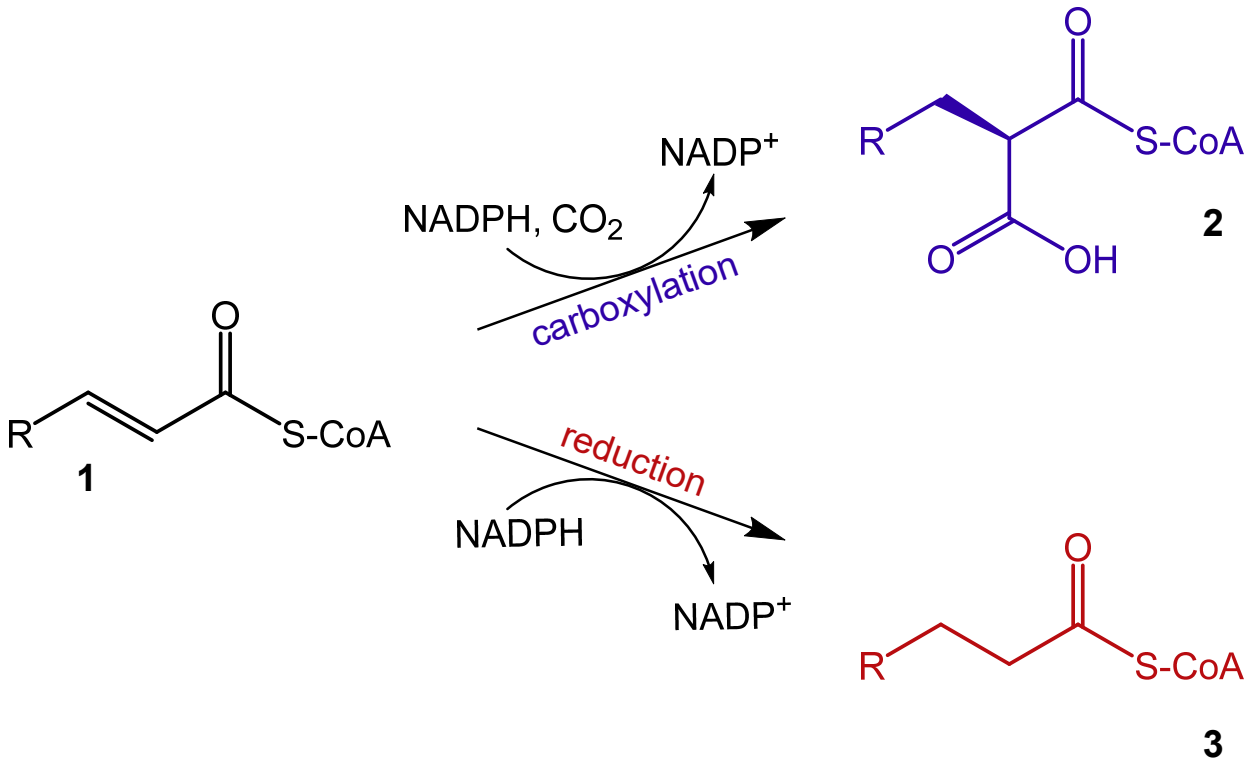


Figure 3: Reaction scheme of enoyl-CoA carboxylase/reductases (ECRs). The ECR can carboxylate the *α,β*-unsaturated substrate **1** NADPH dependently, resulting in **2**, an *α*-substituted malonyl-CoA derivative. In a side reaction or in absence of CO_2 the substrate can be reduced to **3**, an alkyl-CoA. R = alkyl residues.

The genes encoding for secondary ECRs are often located within or in proximity of the biosynthetic gene cluster of the respective PKS that incorporates atypical extender units [44]. After the characterization of their carboxylation activity and role in ethylmalonyl-CoA production, the use of ECRs in synthesizing atypical extender units was soon demonstrated. Not only malonyl-CoA derivatives with linear or branched alkyl-substitutions could be synthesized but also substitutions bearing allyl-groups. These developments led to the production of unprecedented polyketide structures [41, 43, 45-48].

Acyltransferases select starter and extender units

The AT domains are responsible for the selection of specific starter and extender units and consequently also for the *α*-substitutions of the polyketide. Their role in polyketide biosynthesis made them the major target of engineering efforts with the aim of creating novel polyketide structures and introducing new functionalities. ATs generally show a high

specificity for their substrate, which is conveyed by specificity motifs. A common motif lies downstream of the GXSXG sequence motif containing the active site serine. In malonyl-CoA specific AT domains a phenylalanine is placed in proximity to that motif, which is replaced commonly with a glycine or serine in methyl and ethylmalonyl-CoA transacylating ATs [49-51]. Another intensively studied motif is the highly conserved HAFH motif in the binding pocket of malonyl-CoA transacylating AT domains. Methylmalonyl-CoA specific domains mostly present a YASH motif. In AT domains transacylating ethylmalonyl-CoA and longer acyl-residues this motif is less well conserved. In the case of ethylmalonyl-CoA ATs it generally presents the sequence XAGH, with X mainly being F, T, V or H [52]. Attempts have been made to exchange complete sequence motifs or swapping entire AT domains exhibiting different substrate profiles, resulting in the production of chimeric PKS [53-55]. These constructs often resulted in insoluble protein, very low activities and increased promiscuity rather than specificity changes. A less invasive strategy is the exchange of individual specificity conferring amino acids by site-directed mutagenesis, which as well often resulted in promiscuous AT domains [52, 56-58]. The difficulties in achieving full inversion of specificity highlight an incomplete understanding of which amino acid residues are involved in substrate selection. Recently a series of previously unidentified amino acids participating in substrate selection were characterized, where even single mutations led to a shifted selectivity. Guided by the new findings, exchanges of larger structural motifs than anticipated before, led to almost complete shifts towards larger substrates in the pikromycin system [59].

Reports on naturally promiscuous AT domains are increasing, providing platforms for the diversification of polyketides with little to no engineering [60, 61]. The AT5 of monensin PKS selects naturally for methyl- and ethylmalonyl-CoA. Nevertheless, it was shown that this domain also accepts allyl-, propargyl- and butylmalonyl-CoA, resulting in novel monensin derivatives without the need for engineering [62, 63]. The erythromycin modular type I PKS DEBS, generates the macrolide 6-deoxyerythronolide B (6-dEB) from propionyl- and methylmalonyl-CoA. This PKS is the archetypal model system for investigating PKS mechanisms, specificity and engineering. Heterologous production of 6-dEB has been achieved in a metabolically engineered *E.coli* strain BAP1 and later the DEBS system was reconstituted *in vitro* yielding the full macrolide, as well as truncated

lactone versions [64-67]. Initial assumptions that the DEBS AT domains exclusively accept methylmalonyl-CoA as substrate were soon revealed to not hold true, as experiments showed an intrinsic promiscuity leading to the incorporation of butylmalonyl-, propargylmalonyl- or azidoethylmalonyl-CoA[68, 69]. The same was shown for the pikromycin PKS that synthesizes a structurally highly similar polyketide, in which incorporation of, for instance ethyl- and propargylmalonyl-CoA was observed [60, 70]. The inherent promiscuity of AT domains complicates the study of specificity-conferring amino acids and hinders the targeted incorporation of atypical extender units. Yet, at the same time such AT domains offer the opportunity to be exploited as platforms for polyketide diversification.

Thioesterases in polyketide biosynthesis

Once the polyketide chain reaches full length, it is released from its thioester tether of the ACP. Typically, in modular PKS the last biosynthetic module contains a covalently bound, terminal type I thioesterase domain (TEI) which catalyzes the release of the polyketide by hydrolysis or intramolecular macrocyclization. Nonetheless, other release mechanisms are also known [71]. TEs are members of the *α,β* hydrolase superfamily that contain a catalytic Ser-His-Asp triad with the active site serine serving as a nucleophile [72]. Polyketide TEIs are dimeric and contain a hydrophobic substrate channel which stretches through the entire domain to accommodate the polyketide. Release is catalyzed by the transfer of the polyketide chain to the active site serine, forming an acyl-*O*-TE intermediate. This intermediate can either be captured by an external nucleophile like water, or by an internal nucleophile like a hydroxyl- or amine-group leading to a linear or cyclic product, respectively. For cyclization resulting in lactones and lactams the acyl-*O*-TE intermediate must be protected from solvent nucleophiles, while the polyketide chain is precisely positioned in the TE to allow regiospecific capture of the carbonyl by an internal nucleophile [73, 74]. In natural systems the TEI is presented the final, native product. This is usually only a single compound; substrate selection is thus not required. In the case of pikromyin, the terminal TE is releasing the heptaketide as well as a hexaketide [75]. This indicates a relaxed substrate specificity of TEIs. Studies on the terminal TE from DEBS, TE_{DEBS}, and fusions for the production of truncated polyketide derivatives showed

only a modest substrate preference of this TE, favoring the natural (2*R*, 3*S*) configuration of the 2-methyl-3-hydroxy substituents in the presence of a distal hydroxyl-group [76-78].

In addition to the terminal TE, many polyketide biosynthetic clusters contain an external thioesterase, known as type II TE (TEII). The result of knocking out these proteins can reach from no effect, to a complete loss of product formation [79-82]. TEIIs generally have a beneficial effect on increasing the production yield, which can be exploited by co-expression in heterologous systems [83, 84]. Strong overexpression on the other hand is observed to reduce product formation [82, 85]. The main role of TEIIs is suggested to be the removal of modifications blocking the reactive thiol of the 4'-phosphohantetheinyl arm of ACPs, leading to a stalling of the biosynthesis. Besides this clearing activity, also other functions have been observed [80, 81, 86-90]. Two potential origins of structures blocking the assembly line are discussed in literature. The first is decarboxylation of extender units tethered to the ACP by the KS domain, without following elongation [91-93]. The second is the conversion of the *apo*-ACP into the *holo*-ACP by phosphopantetheinyl transferases that utilize acyl-CoA thioesters instead of free CoA-SH. This leads to the loading of the acyl residue onto the ACP and consequently to a blockage of the assembly line.

Furthermore, two models for the action of TEII have been proposed. In the high specificity model, the TEII efficiently removes exclusively aberrant acyl residues. In the low specificity model, the TEII inefficiently removes correct and aberrant acyl residues [91, 94]. The correct residues, including growing polyketide chains are protected from hydrolysis by efficient processing through the PKS, aberrant residues cannot be processed, or only slowly, and are exposed on the surface for an increased period of time, enabling the TEII to attack and hydrolyze the residues.

Aim of this study

After recognizing the native incorporation of various *α*-substituted extender units into polyketides, investigations of AT domains aimed to help understand the biochemistry behind substrate selectivity. Inspired by this, the production of novel polyketides mediated by the incorporation of non-native extender units was attempted through assembly line engineering. A major difficulty that limits thorough AT characterization, is the availability of extender units. The commonly used synthetic production route of extender units is the

utilization of a promiscuous MatB variant [68, 95, 96]. This route however shows many disadvantages. **Chapter I** describes how ECRs can be used for the production of a versatile set of *α*-substituted malonyl-CoA derivatives. The carboxylation efficiency of a promiscuous ECR was maximized by coupling the reaction to a biocatalytic proofreading mechanism. For this, an oxidase was employed that recycles the reduction side reaction products of the ECR back into the *α*,*β* unsaturated starting substrates. Furthermore, a convenient preparative scale, high purity assuring “one-pot” workflow for the production of atypical extender units was established and the incorporation of these extender units into polyketides could be demonstrated. With this study, the basis to facilitate the exploration of AT domains and PKS promiscuity and engineering is set.

Only a few reliable and well established PKS model systems are available for *in vitro* studies. Two of them are the archetypal modular DEBS system and the PikA PKS. Both of these systems show intrinsic promiscuity of multiple AT domains [60, 68, 70, 97], which complicates the study of selectivity conferring properties. To fundamentally understand how extender unit selectivity is achieved, more diverse systems are required. **Chapter II** focuses on the modular type I PKS of phoslactomycin (Pn PKS) from *Streptomyces platensis*. This PKS system displays unique features, like the selection for the rare starter unit cyclohexanecarboxyl-CoA, the incorporation of five molecules malonyl- and two molecules ethylmalonyl-CoA and the generation of two *cis*-double bonds. These properties make Pn PKS an excellent model to study selectivity at different sites in one PKS. This chapter describes the successful *in vitro* reconstitution of the first six modules of Pn PKS. Competitive substrate incorporation assays revealed a module with high substrate tolerance. Combined with a steady state kinetic assay, established for this work, the reaction parameters of excised AT domains could be assessed. With this, the mechanism behind substrate selection of this promiscuous AT domain could be unraveled. This study provides a PKS model system, complementary to those commonly used, and demonstrates the pursue of mechanistic questions in the context of an under-studied PKS.

PKS *in vitro* and *in vivo* studies often neglect possible effects on the polyketide production yields, which could be caused by additional enzymes present in the producer organisms. Most information about the function of type II thioesterases, enzymes encoded in the majority of polyketide biosynthetic gene clusters, has been gained by *in vivo* knock out experiments. Only a few *in vitro* experiments have been done combining TEIIs with their associated biosynthetic enzymes [85, 98], while the impact of TEIIs on polyketide production of modular type I PKS *in vitro* remains unexplored. In **Chapter III**, the effects of PnG, the TEII from the phoslactomycin PKS, on the previously reconstituted *in vitro* system are presented. Kinetic analysis of PnG with ACP tethered (alkyl)malonyl- and alkyl- residues showed a strong preference for alkyl-ACPs, representing assembly line blocking modifications. When added to the *in vitro* Pn PKS system, low amounts of PnG increased polyketide production of the assembly line, suggesting substantial blocking thereof. PnG could moreover be shown to release polyketides, functionally replacing the terminal thioesterase *in vitro*. Careful evaluation of the data suggests that the origin of the residues blocking the assembly line are the miss-loading by phosphopantetheinyl transferases but as well as decarboxylation of extender units *in situ*. Put together, the results support the low specificity model of TEIIs and highlight the utility of these enzymes for improving natural product biosynthesis *in vitro*.

References

1. Newman, D. J.; Cragg, G. M.; Snader, K. M., The influence of natural products upon drug discovery. *Natural product reports* **2000,** *17* (3), 215-234.
2. Hertweck, C., The biosynthetic logic of polyketide diversity. *Angew Chem Int Ed Engl* **2009,** *48* (26), 4688-716.
3. Staunton, J.; Weissman, K. J., Polyketide biosynthesis: a millennium review. *Natural product reports* **2001,** *18* (4), 380-416.
4. Katz, L.; Baltz, R. H., Natural product discovery: past, present, and future. *Journal of industrial microbiology & biotechnology* **2016,** *43* (2-3), 155-176.
5. Newman, D. J.; Cragg, G. M., Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *Journal of Natural Products* **2020,** *83* (3), 770-803.
6. Cooke, T. F.; Fischer, C. R.; Wu, P.; Jiang, T.-X.; Xie, K. T.; Kuo, J.; Doctorov, E.; Zehnder, A.; Khosla, C.; Chuong, C.-M., Genetic mapping and biochemical basis of yellow feather pigmentation in budgerigars. *Cell* **2017,** *171* (2), 427-439. e21.
7. Barka, E. A.; Vatsa, P.; Sanchez, L.; Gaveau-Vaillant, N.; Jacquard, C.; Klenk, H.-P.; Clément, C.; Ouhdouch, Y.; van Wezel, G. P., Taxonomy, physiology, and natural

products of Actinobacteria. *Microbiology and Molecular Biology Reviews* **2016,** *80* (1), 1-43.
8. Kondo, S.; Yasui, K.; Natsume, M.; Katayama, M.; Marumo, S., Isolation, physico-chemical properties and biological activity of pamamycin-607, an aerial mycelium-inducing substance from Streptomyces alboniger. *The Journal of antibiotics* **1988,** *41* (9), 1196-1204.
9. Jez, J. M.; Austin, M. B.; Ferrer, J.-L.; Bowman, M. E.; Schröder, J.; Noel, J. P., Structural control of polyketide formation in plant-specific polyketide synthases. *Chemistry & biology* **2000,** *7* (12), 919-930.
10. Malico, A. A.; Nichols, L.; Williams, G. J., Synthetic biology enabling access to designer polyketides. *Current Opinion in Chemical Biology* **2020,** *58*, 45-53.
11. Cortes, J.; Haydock, S. F.; Roberts, G. A.; Bevitt, D. J.; Leadlay, P. F., An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of Saccharopolyspora erythraea. *Nature* **1990,** *348* (6297), 176-178.
12. Hopwood, D. A.; Sherman, D. H., Molecular genetics of polyketides and its comparison to fatty acid biosynthesis. *Annual review of genetics* **1990,** *24* (1), 37-62.
13. Cox, R. J., Polyketides, proteins and genes in fungi: programmed nano-machines begin to reveal their secrets. *Organic & biomolecular chemistry* **2007,** *5* (13), 2010-2026.
14. Donadio, S.; Staver, M. J.; McAlpine, J. B.; Swanson, S. J.; Katz, L., Modular organization of genes required for complex polyketide biosynthesis. *Science* **1991,** *252* (5006), 675-679.
15. Moss, S. J.; Martin, C. J.; Wilkinson, B., Loss of co-linearity by modular polyketide synthases: a mechanism for the evolution of chemical diversity. *Natural product reports* **2004,** *21* (5), 575-593.
16. Piel, J.; Wen, G.; Platzer, M.; Hui, D., Unprecedented diversity of catalytic domains in the first four modules of the putative pederin polyketide synthase. *Chembiochem : a European journal of chemical biology* **2004,** *5* (1), 93-98.
17. Helfrich, E. J.; Ueoka, R.; Dolev, A.; Rust, M.; Meoded, R. A.; Bhushan, A.; Califano, G.; Costa, R.; Gugger, M.; Steinbeck, C., Automated structure prediction of trans-acyltransferase polyketide synthase products. *Nature chemical biology* **2019,** *15* (8), 813-821.
18. Hertweck, C.; Luzhetskyy, A.; Rebets, Y.; Bechthold, A., Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. *Natural product reports* **2007,** *24* (1), 162-190.
19. Austin, M. B.; Noel, J. P., The chalcone synthase superfamily of type III polyketide synthases. *Natural product reports* **2003,** *20* (1), 79-110.
20. Dutta, S.; Whicher, J. R.; Hansen, D. A.; Hale, W. A.; Chemler, J. A.; Congdon, G. R.; Narayan, A. R.; Håkansson, K.; Sherman, D. H.; Smith, J. L., Structure of a modular polyketide synthase. *Nature* **2014,** *510* (7506), 512-517.
21. Tang, Y.; Kim, C.-Y.; Mathews, I. I.; Cane, D. E.; Khosla, C., The 2.7-Å crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proceedings of the National Academy of Sciences* **2006,** *103* (30), 11124-11129.
22. Zheng, J.; Gay, D. C.; Demeler, B.; White, M. A.; Keatinge-Clay, A. T., Divergence of multimodular polyketide synthases revealed by a didomain structure. *Nature chemical biology* **2012,** *8* (7), 615-21.

23. Khare, D.; Hale, W. A.; Tripathi, A.; Gu, L.; Sherman, D. H.; Gerwick, W. H.; Håkansson, K.; Smith, J. L., Structural Basis for Cyclopropanation by a Unique Enoyl-Acyl Carrier Protein Reductase. *Structure* **2015,** *23* (12), 2213-2223.
24. Keatinge-Clay, A., Crystal structure of the erythromycin polyketide synthase dehydratase. *J Mol Biol* **2008,** *384* (4), 941-53.
25. Keatinge-Clay, A. T.; Stroud, R. M., The structure of a ketoreductase determines the organization of the beta-carbon processing enzymes of modular polyketide synthases. *Structure* **2006,** *14* (4), 737-48.
26. Tang, Y.; Chen, A. Y.; Kim, C.-Y.; Cane, D. E.; Khosla, C., Structural and mechanistic analysis of protein interactions in module 3 of the 6-deoxyerythronolide B synthase. *Chemistry & biology* **2007,** *14* (8), 931-943.
27. Klaus, M.; Grininger, M., Engineering strategies for rational polyketide synthase design. *Natural product reports* **2018,** *35* (10), 1070-1081.
28. Dodge, G. J.; Maloney, F. P.; Smith, J. L., Protein–protein interactions in “cis-AT” polyketide synthases. *Natural product reports* **2018,** *35* (10), 1082-1096.
29. Maier, T.; Leibundgut, M.; Ban, N., The crystal structure of a mammalian fatty acid synthase. *Science* **2008,** *321* (5894), 1315-1322.
30. Zheng, J.; Fage, C. D.; Demeler, B.; Hoffman, D. W.; Keatinge-Clay, A. T., The missing linker: a dimerization motif located within polyketide synthase modules. *ACS chemical biology* **2013,** *8* (6), 1263-70.
31. Broadhurst, R. W.; Nietlispach, D.; Wheatcroft, M. P.; Leadlay, P. F.; Weissman, K. J., The structure of docking domains in modular polyketide synthases. *Chemistry & biology* **2003,** *10* (8), 723-31.
32. Herbst, D. A.; Jakob, R. P.; Zähringer, F.; Maier, T., Mycocerosic acid synthase exemplifies the architecture of reducing polyketide synthases. *Nature* **2016,** *531* (7595), 533-7.
33. Whicher, J. R.; Dutta, S.; Hansen, D. A.; Hale, W. A.; Chemler, J. A.; Dosey, A. M.; Narayan, A. R.; Håkansson, K.; Sherman, D. H.; Smith, J. L.; Skiniotis, G., Structural rearrangements of a polyketide synthase module during its catalytic cycle. *Nature* **2014,** *510* (7506), 560-4.
34. Chen, Y. L.; Zhao, J.; Liu, W.; Gao, J. F.; Tao, L. M.; Pan, H. X.; Tang, G. L., Identification of phoslactomycin biosynthetic gene clusters from Streptomyces platensis SAM-0654 and characterization of PnR1 and PnR2 as positive transcriptional regulators. *Gene* **2012,** *509* (2), 195-200.
35. Ikeda, H.; Ōmura, S., Avermectin biosynthesis. *Chemical Reviews* **1997,** *97* (7), 2591-2610.
36. Moore, B. S.; Hertweck, C., Biosynthesis and attachment of novel bacterial polyketide synthase starter units. *Natural product reports* **2002,** *19* (1), 70-99.
37. Wongkittichote, P.; Mew, N. A.; Chapman, K. A., Propionyl-CoA carboxylase–a review. *Molecular genetics and metabolism* **2017,** *122* (4), 145-152.
38. Chang, C.; Huang, R.; Yan, Y.; Ma, H.; Dai, Z.; Zhang, B.; Deng, Z.; Liu, W.; Qu, X., Uncovering the formation and selection of benzylmalonyl-CoA from the biosynthesis of splenocin and enterocin reveals a versatile way to introduce amino acids into polyketide carbon scaffolds. *Journal of the American Chemical Society* **2015,** *137* (12), 4183-4190.

39. Quade, N.; Huo, L.; Rachid, S.; Heinz, D. W.; Müller, R., Unusual carbon fixation gives rise to diverse polyketide extender units. *Nature chemical biology* **2012,** *8* (1), 117-124.
40. Yoo, H.-G.; Kwon, S.-Y.; Kim, S.; Karki, S.; Park, Z.-Y.; Kwon, H.-J., Characterization of 2-octenoyl-CoA carboxylase/reductase utilizing pteB from Streptomyce avermitilis. *Bioscience, biotechnology, and biochemistry* **2011**, 1105102493-1105102493.
41. Mo, S.; Kim, D. H.; Lee, J. H.; Park, J. W.; Basnet, D. B.; Ban, Y. H.; Yoo, Y. J.; Chen, S.-w.; Park, S. R.; Choi, E. A., Biosynthesis of the allylmalonyl-CoA extender unit for the FK506 polyketide synthase proceeds through a dedicated polyketide synthase and facilitates the mutasynthesis of analogues. *Journal of the American Chemical Society* **2011,** *133* (4), 976-985.
42. Erb, T. J., Carboxylases in natural and synthetic microbial pathways. *Applied and environmental microbiology* **2011,** *77* (24), 8466-8477.
43. Erb, T. J.; Berg, I. A.; Brecht, V.; Müller, M.; Fuchs, G.; Alber, B. E., Synthesis of C5-dicarboxylic acids from C2-units involving crotonyl-CoA carboxylase/reductase: the ethylmalonyl-CoA pathway. *Proceedings of the National Academy of Sciences* **2007,** *104* (25), 10631-10636.
44. Wilson, M. C.; Moore, B. S., Beyond ethylmalonyl-CoA: the functional role of crotonyl-CoA carboxylase/reductase homologs in expanding polyketide diversity. *Natural product reports* **2012,** *29* (1), 72-86.
45. Peter, D. M.; Schada von Borzyskowski, L.; Kiefer, P.; Christen, P.; Vorholt, J. A.; Erb, T. J., Screening and engineering the synthetic potential of carboxylating reductases from central metabolism and polyketide biosynthesis. *Angewandte Chemie International Edition* **2015,** *54* (45), 13457-13461.
46. Erb, T. J.; Brecht, V.; Fuchs, G.; Müller, M.; Alber, B. E., Carboxylation mechanism and stereochemistry of crotonyl-CoA carboxylase/reductase, a carboxylating enoyl-thioester reductase. *Proceedings of the National Academy of Sciences* **2009,** *106* (22), 8871-8876.
47. Liu, Y.; Hazzard, C.; Eustáquio, A. S.; Reynolds, K. A.; Moore, B. S., Biosynthesis of salinosporamides from α, β-unsaturated fatty acids: implications for extending polyketide synthase diversity. *Journal of the American Chemical Society* **2009,** *131* (30), 10376-10377.
48. Yan, Y.; Chen, J.; Zhang, L.; Zheng, Q.; Han, Y.; Zhang, H.; Zhang, D.; Awakawa, T.; Abe, I.; Liu, W., Multiplexing of combinatorial chemistry in antimycin biosynthesis: expansion of molecular diversity and utility. *Angewandte Chemie* **2013,** *125* (47), 12534-12538.
49. Reeves, C. D.; Murli, S.; Ashley, G. W.; Piagentini, M.; Hutchinson, C. R.; McDaniel, R., Alteration of the substrate specificity of a modular polyketide synthase acyltransferase domain through site-specific mutations. *Biochemistry* **2001,** *40* (51), 15464-70.
50. Khayatt, B. I.; Overmars, L.; Siezen, R. J.; Francke, C., Classification of the adenylation and acyl-transferase activity of NRPS and PKS systems using ensembles of substrate specific hidden Markov models. *PLoS One* **2013,** *8* (4), e62136.

51. Yadav, G.; Gokhale, R. S.; Mohanty, D., Computational approach for prediction of domain organization and substrate specificity of modular polyketide synthases. *Journal of molecular biology* **2003,** *328* (2), 335-363.
52. Del Vecchio, F.; Petkovic, H.; Kendrew, S. G.; Low, L.; Wilkinson, B.; Lill, R.; Cortes, J.; Rudd, B. A.; Staunton, J.; Leadlay, P. F., Active-site residue, domain and module swaps in modular polyketide synthases. *J Ind Microbiol Biotechnol* **2003,** *30* (8), 489-94.
53. Oliynyk, M.; Brown, M. J.; Cortes, J.; Staunton, J.; Leadlay, P. F., A hybrid modular polyketide synthase obtained by domain swapping. *Chemistry & biology* **1996,** *3* (10), 833-9.
54. McDaniel, R.; Thamchaipenet, A.; Gustafsson, C.; Fu, H.; Betlach, M.; Ashley, G., Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products. *Proceedings of the National Academy of Sciences of the United States of America* **1999,** *96* (5), 1846-51.
55. Lau, J.; Fu, H.; Cane, D. E.; Khosla, C., Dissecting the role of acyltransferase domains of modular polyketide synthases in the choice and stereochemical fate of extender units. *Biochemistry* **1999,** *38* (5), 1643-51.
56. Sundermann, U.; Bravo-Rodriguez, K.; Klopries, S.; Kushnir, S.; Gomez, H.; Sanchez-Garcia, E.; Schulz, F., Enzyme-directed mutasynthesis: a combined experimental and theoretical approach to substrate recognition of a polyketide synthase. *ACS chemical biology* **2013,** *8* (2), 443-50.
57. Bravo-Rodriguez, K.; Klopries, S.; Koopmans, K. R. M.; Sundermann, U.; Yahiaoui, S.; Arens, J.; Kushnir, S.; Schulz, F.; Sanchez-Garcia, E., Substrate Flexibility of a Mutated Acyltransferase Domain and Implications for Polyketide Biosynthesis. *Chemistry & biology* **2015,** *22* (11), 1425-1430.
58. Koryakina, I.; Kasey, C.; McArthur, J. B.; Lowell, A. N.; Chemler, J. A.; Li, S.; Hansen, D. A., Inversion of Extender Unit Selectivity in the Erythromycin Polyketide Synthase by Acyltransferase Domain Engineering. **2017,** *12* (1), 114-123.
59. Kalkreuter, E.; Bingham, K. S.; Keeler, A. M.; Lowell, A.; Schmidt, J. J.; Sherman, D. H.; Williams, G. J., Computationally-guided exchange of substrate selectivity motifs in a modular polyketide synthase acyltransferase. *bioRxiv* **2020**.
60. Kalkreuter, E.; CroweTipton, J. M.; Lowell, A. N., Engineering the Substrate Specificity of a Modular Polyketide Synthase for Installation of Consecutive Non-Natural Extender Units. **2019,** *141* (5), 1961-1969.
61. Geyer, K.; Sundaram, S.; Susnik, P.; Koert, U.; Erb, T. J., Understanding Substrate Selectivity of Phoslactomycin Polyketide Synthase by Using Reconstituted in Vitro Systems. *Chembiochem : a European journal of chemical biology* **2020**.
62. Oliynyk, M.; Stark, C. B.; Bhatt, A.; Jones, M. A.; Hughes-Thomas, Z. A.; Wilkinson, C.; Oliynyk, Z.; Demydchuk, Y.; Staunton, J.; Leadlay, P. F., Analysis of the biosynthetic gene cluster for the polyether antibiotic monensin in Streptomyces cinnamonensis and evidence for the role of monB and monC genes in oxidative cyclization. *Molecular microbiology* **2003,** *49* (5), 1179-90.
63. Ismail-Ali, A.; Fansa, E. K.; Pryk, N.; Yahiaoui, S.; Kushnir, S.; Pflieger, M.; Wittinghofer, A.; Schulz, F., Biosynthesis-driven structure-activity relationship study of premonensin-derivatives. *Org Biomol Chem* **2016,** *14* (32), 7671-5.

64. Pfeifer, B. A.; Admiraal, S. J.; Gramajo, H.; Cane, D. E.; Khosla, C., Biosynthesis of complex polyketides in a metabolically engineered strain of E. coli. *Science* **2001,** *291* (5509), 1790-2.
65. Pieper, R.; Luo, G.; Cane, D. E.; Khosla, C., Cell-free synthesis of polyketides by recombinant erythromycin polyketide synthases. *Nature* **1995,** *378* (6554), 263-6.
66. Marsden, A. F.; Caffrey, P.; Aparicio, J. F.; Loughran, M. S.; Staunton, J.; Leadlay, P. F., Stereospecific acyl transfers on the erythromycin-producing polyketide synthase. *Science* **1994,** *263* (5145), 378-80.
67. Kao, C. M.; Luo, G.; Katz, L.; Cane, D. E.; Khosla, C., Manipulation of macrolide ring size by directed mutagenesis of a modular polyketide synthase. *Journal of the American Chemical Society* **1995,** *117* (35), 9105-9106.
68. Koryakina, I.; McArthur, J. B.; Draelos, M. M.; Williams, G. J., Promiscuity of a modular polyketide synthase towards natural and non-natural extender units. *Org Biomol Chem* **2013,** *11* (27), 4449-58.
69. Pohl, N. L.; Hans, M.; Lee, H. Y.; Kim, Y. S.; Cane, D. E.; Khosla, C., Remarkably broad substrate tolerance of malonyl-CoA synthetase, an enzyme capable of intracellular synthesis of polyketide precursors. *J Am Chem Soc* **2001,** *123* (24), 5822-3.
70. Bonnett, S. A.; Rath, C. M.; Shareef, A. R.; Joels, J. R.; Chemler, J. A.; Hakansson, K.; Reynolds, K.; Sherman, D. H., Acyl-CoA subunit selectivity in the pikromycin polyketide synthase PikAIV: steady-state kinetics and active-site occupancy analysis by FTICR-MS. *Chemistry & biology* **2011,** *18* (9), 1075-81.
71. Du, L.; Lou, L., PKS and NRPS release mechanisms. *Natural product reports* **2010,** *27* (2), 255-78.
72. Nardini, M.; Dijkstra, B. W., α/β Hydrolase fold enzymes: the family keeps growing. *Current opinion in structural biology* **1999,** *9* (6), 732-737.
73. Kohli, R. M.; Walsh, C. T., Enzymology of acyl chain macrocyclization in natural product biosynthesis. *Chemical communications* **2003,** (3), 297-307.
74. Keating, T. A.; Walsh, C. T., Initiation, elongation, and termination strategies in polyketide and polypeptide antibiotic biosynthesis. *Current opinion in chemical biology* **1999,** *3* (5), 598-606.
75. Akey, D. L.; Kittendorf, J. D.; Giraldes, J. W.; Fecik, R. A.; Sherman, D. H.; Smith, J. L., Structural basis for macrolactonization by the pikromycin thioesterase. *Nature chemical biology* **2006,** *2* (10), 537-542.
76. Gokhale, R. S.; Hunziker, D.; Cane, D. E.; Khosla, C., Mechanism and specificity of the terminal thioesterase domain from the erythromycin polyketide synthase. *Chemistry & biology* **1999,** *6* (2), 117-125.
77. Weissman, K. J.; Smith, C. J.; Hanefeld, U.; Aggarwal, R.; Bycroft, M.; Staunton, J.; Leadlay, P. F., The Thioesterase of the Erythromycin-Producing Polyketide Synthase: Influence of Acyl Chain Structure on the Mode of Release of Substrate Analogues from the Acyl Enzyme Intermediates. *Angew Chem Int Ed Engl* **1998,** *37* (10), 1437-1440.
78. Aggarwal, R.; Caffrey, P.; Leadlay, P. F.; Smith, C. J.; Staunton, J., The thioesterase of the erythromycin-producing polyketide synthase: mechanistic studies in vitro to investigate its mode of action and substrate specificity. *Journal of the Chemical Society, Chemical Communications* **1995,** (15), 1519-1520.

79. Kotowska, M.; Pawlik, K.; Butler, A. R.; Cundliffe, E.; Takano, E.; Kuczek, K., Type II thioesterase from Streptomyces coelicolor A3(2). *Microbiology (Reading, England)* **2002,** *148* (Pt 6), 1777-1783.
80. Wu, H.; Liang, J.; Gou, L.; Wu, Q.; Liang, W. J.; Zhou, X.; Bruce, I. J.; Deng, Z.; Wang, Z., Recycling of Overactivated Acyls by a Type II Thioesterase during Calcimycin Biosynthesis in Streptomyces chartreusis NRRL 3882. *Applied and environmental microbiology* **2018,** *84* (12).
81. Butler, A. R.; Bate, N.; Cundliffe, E., Impact of thioesterase activity on tylosin biosynthesis in Streptomyces fradiae. *Chemistry & biology* **1999,** *6* (5), 287-92.
82. Kim, B. S.; Cropp, T. A.; Beck, B. J.; Sherman, D. H.; Reynolds, K. A., Biochemical evidence for an editing role of thioesterase II in the biosynthesis of the polyketide pikromycin. *The Journal of biological chemistry* **2002,** *277* (50), 48028-34.
83. Kotowska, M.; Pawlik, K., Roles of type II thioesterases and their application for secondary metabolite yield improvement. *Applied microbiology and biotechnology* **2014,** *98* (18), 7735-46.
84. Pfeifer, B.; Hu, Z.; Licari, P.; Khosla, C., Process and metabolic strategies for improved production of Escherichia coli-derived 6-deoxyerythronolide B. *Applied and environmental microbiology* **2002,** *68* (7), 3287-92.
85. Ohlemacher, S. I.; Xu, Y.; Kober, D. L.; Malik, M.; Nix, J. C.; Brett, T. J.; Henderson, J. P., YbtT is a low-specificity type II thioesterase that maintains production of the metallophore yersiniabactin in pathogenic enterobacteria. *The Journal of biological chemistry* **2018,** *293* (51), 19572-19585.
86. Schneider, A.; Marahiel, M. A., Genetic evidence for a role of thioesterase domains, integrated in or associated with peptide synthetases, in non-ribosomal peptide biosynthesis in Bacillus subtilis. *Archives of microbiology* **1998,** *169* (5), 404-10.
87. Harvey, B. M.; Hong, H.; Jones, M. A.; Hughes-Thomas, Z. A.; Goss, R. M.; Heathcote, M. L.; Bolanos-Garcia, V. M.; Kroutil, W.; Staunton, J.; Leadlay, P. F.; Spencer, J. B., Evidence that a novel thioesterase is responsible for polyketide chain release during biosynthesis of the polyether ionophore monensin. *Chembiochem : a European journal of chemical biology* **2006,** *7* (9), 1435-42.
88. Liu, T.; You, D.; Valenzano, C.; Sun, Y.; Li, J.; Yu, Q.; Zhou, X.; Cane, D. E.; Deng, Z., Identification of NanE as the thioesterase for polyether chain release in nanchangmycin biosynthesis. *Chemistry & biology* **2006,** *13* (9), 945-55.
89. Kalaitzis, J. A.; Cheng, Q.; Meluzzi, D.; Xiang, L.; Izumikawa, M.; Dorrestein, P. C.; Moore, B. S. J. B.; chemistry, m., Policing starter unit selection of the enterocin type II polyketide synthase by the type II thioesterase EncL. **2011,** *19* (22), 6633-6638.
90. Tang, Y.; Koppisch, A. T.; Khosla, C., The acyltransferase homologue from the initiation module of the R1128 polyketide synthase is an acyl-ACP thioesterase that edits acetyl primer units. *Biochemistry* **2004,** *43* (29), 9546-55.
91. Heathcote, M. L.; Staunton, J.; Leadlay, P. F., Role of type II thioesterases: evidence for removal of short acyl chains produced by aberrant decarboxylation of chain extender units. *Chemistry & biology* **2001,** *8* (2), 207-20.
92. Pieper, R.; Ebert-Khosla, S.; Cane, D.; Khosla, C., Erythromycin biosynthesis: kinetic studies on a fully active modular polyketide synthase using natural and unnatural substrates. *Biochemistry* **1996,** *35* (7), 2054-60.

93. Jacobsen, J. R.; Cane, D. E.; Khosla, C., Spontaneous priming of a downstream module in 6-deoxyerythronolide B synthase leads to polyketide biosynthesis. *Biochemistry* **1998,** *37* (14), 4928-34.
94. Claxton, H. B.; Akey, D. L.; Silver, M. K.; Admiraal, S. J.; Smith, J. L., Structure and functional analysis of RifR, the type II thioesterase from the rifamycin biosynthetic pathway. *The Journal of biological chemistry* **2009,** *284* (8), 5021-9.
95. Koryakina, I.; McArthur, J.; Randall, S.; Draelos, M. M.; Musiol, E. M.; Muddiman, D. C.; Weber, T.; Williams, G. J., Poly specific trans-acyltransferase machinery revealed via engineered acyl-CoA synthetases. *ACS chemical biology* **2013,** *8* (1), 200-8.
96. Koryakina, I.; Williams, G. J., Mutant malonyl-CoA synthetases with altered specificity for polyketide synthase extender unit generation. *Chembiochem : a European journal of chemical biology* **2011,** *12* (15), 2289-93.
97. Lowry, B.; Robbins, T.; Weng, C. H.; O'Brien, R. V.; Cane, D. E.; Khosla, C., In vitro reconstitution and analysis of the 6-deoxyerythronolide B synthase. *J Am Chem Soc* **2013,** *135* (45), 16809-12.
98. Miller, D. A.; Luo, L.; Hillson, N.; Keating, T. A.; Walsh, C. T., Yersiniabactin synthetase: a four-protein assembly line producing the nonribosomal peptide/polyketide hybrid siderophore of Yersinia pestis. *Chemistry & biology* **2002,** *9* (3), 333-44.

Chapter I

Combining promiscuous acyl-CoA oxidase and enoyl-CoA carboxylase/reductases for atypical polyketide extender unit biosynthesis

Published in: Cell Chemical Biology 2018, 25 (7), 833-839

Combining promiscuous acyl-CoA oxidase and enoyl-CoA carboxylase/reductases for atypical polyketide extender unit biosynthesis

Bastian Vögeli[a], Kyra Geyer[a], Patrick D. Gerlinger[a], Sarah Benkstein[a], Niña Socorro Cortina[a], Tobias J. Erb*[a]

[a] Biochemistry and Synthetic Metabolism, Max-Planck-Institute for terrestrial Microbiology, Karl-von-Frisch-Strasse 10, 35043 Marburg (Germany); * Corresponding author: toerb@mpi-marburg.mpg.de

Cell Chemical Biology 2018, 25 (7), 833-839

Author contributions: B.V. and T.J.E conceived the project. B.V., P.D.G., K.G., S.B., and T.J.E. designed and performed experiments and analyzed the data. K.G performed the kinetic measurements and carboxylation efficiencies. P.D.G. conducted the competition assays. P.D.G. and N.S.C. performed mass spectrometry and analysis of the triketide synthesis. B.V., K.G., P.D.G., and T.J.E. wrote the manuscript with contributions from all authors.

Summary

The incorporation of different extender units generates structural diversity in polyketides. There is significant interest in engineering substrate specificity of polyketide synthases (PKSs) to change their chemical structure. Efforts to change extender unit selectivity are hindered by the lack of simple screening methods and easily available atypical extender units. Here, we present a chemo-biosynthetic strategy that employs biocatalytic proofreading and allows access to a large variety of extender units. First, saturated acids are chemically coupled to free coenzyme A (CoA). The corresponding acyl-CoAs are then converted to alkyl-malonyl-CoAs in a ‘‘one-pot’’ reaction through the combined action of an acyl-CoA oxidase and enoyl-CoA carboxylase/reductase. We synthesized six different extender units and used them in *in vitro* competition screens to investigate active site residues conferring extender unit selectivity. Our results show the importance of an uncharacterized glutamine in extender unit selectivity and open the possibility for comprehensive studies on extender incorporation in PKSs.

Introduction

Polyketides are a large class of structurally diverse natural products that exhibit a myriad of biological activities ranging from antibacterial, to antifungal, cholesterol lowering, immune system suppressing, and more [1]. Although polyketides show a large structural and functional variety, they are all assembled from simple acyl- and alkylmalonyl-thioester building blocks. The building blocks are combined in successive Claisen-condensation reactions catalyzed by large multidomain enzymes, so-called polyketide synthases (PKSs). Structural diversity in the final polyketide product is achieved by a variety of different mechanisms. One important mechanism to generate structural variety in polyketides is the use of different acyl-coenzyme A (CoA) building blocks ("starter units") and their condensation with different alkymalonyl-CoA building blocks ("extender units") during biosynthesis [2]. While PKSs have been recognized to accept a variety of different acyl-CoA starter units [3] , the extension of the growing polyketide chain was for a long time believed to be restricted to malonyl-CoA and methylmalonyl-CoA units. However, the growing number of PKSs that are able to incorporate atypical alkylmalonyl-CoA extender units, such as ethylmalonyl-, benzylmalonyl-, or octanoylmalonyl-CoA, highlights the importance of this principle in the generation of structurally and functionally diverse polyketides [2].

Most of the atypical extender units are provided via the reductive carboxylation of unsaturated acyl-CoA precursors by enoyl-CoA carboxylase/reductases [4-6]. Genes encoding carboxylating enoyl-thioester reductases (ECRs) are often associated with the biosynthetic gene cluster of the respective PKS that incorporates the atypical extender unit. Recent studies on the biochemistry of the ECRs identified active site residues that determine the substrate spectrum of these enzymes [7]. According to these studies, ECRs can be categorized into two subfamilies; the ECR-1 subfamily that possess a rather narrow substrate spectrum, and the ECR-2 subfamily that display a more relaxed substrate specificity [6, 7].

The incorporation of atypical extender units through the targeted reprogramming of PKSs has become an important focus in bioengineering to create polyketides that show novel properties. Most approaches in this direction focus on the manipulation of the acyltransferase (AT) domains that are part of the PKS biosynthetic machinery. ATs are

often described as the “gatekeeper” domains that decide which extender units will enter the PKS assembly line. In contrast, the downstream domains of the PKSs that perform the actual condensation and modification reactions are believed to be rather promiscuous towards unnatural *α*-substituents [8-11]. Several different strategies to manipulate the specificity of extender unit incorporation have been applied up to date [12-14]. Although it has become possible to incorporate atypical extender units, this was mainly achieved by lowering selectivity of ATs towards their native extender unit, which is very often accompanied by strongly reduced overall activities. An improved specificity by improving the desired selectivity in combination with reducing the native selectivity by a point mutation has to our knowledge only been achieved in one case [13]. This highlights the necessity for more comprehensive screening methods to fully understand the contributions of ATs and other PKS domains toward extender unit specificity.

Another difficulty that complicates AT engineering is the limited availability of atypical extender unit variants that can be used to screen for altered specificities of ATs. Commonly atypical extender units are synthesized *in situ* via enzymatic coupling of malonic acid derivatives and CoA using a promiscuous variant of the malonyl-CoA synthetase MatB from *Rhizobium trifolii* [10, 11, 15]. This synthesis route is, however, restricted by the limited number of commercially available malonic acid derivatives, which otherwise have to be laboriously synthesized via the acylation of Meldrum's acid [16]. The recent description of promiscuous ECRs has opened up new routes for the chemo-enzymatic synthesis of atypical extender units at preparative scale[7]. However, promiscuous ECRs often display a significant side reactivity and simply reduce the unsaturated acyl-CoA precursors instead of carboxylation. As an example, AntE, the reductive carboxylase in the antimycin biosynthesis pathway, carboxylates only 25% of one of its natural substrates cinnamoyl-CoA, while the rest of the substrate is reduced to phenylpropionyl-CoA (“dihydrocinnamoyl-CoA”) [17].

Here we increased the carboxylation yield of promiscuous ECRs by applying the principle of biocatalytic proofreading. We successfully coupled promiscuous ECRs with an acyl-CoA oxidase that is able to continuously recycle the reduction reaction side product, pushing the ECR reaction toward the carboxylation product. Based on this principle, we further developed a chemo-biosynthetic route that allows for the convenient

preparation of atypical alkylmalonyl-CoA extender units from simple starting materials in “one pot” and at high carboxylation yield. We use this route to prepare a set of six different alkyl-mlalonyl-CoA extender units at preparative scale and high purity, and demonstrate with a recently established *in vitro* model system of the DEBS PKS [18] that these extender units can be used to systematically screen the extender unit selectivity of PKSs.

Results and Discussion

Screening of ECRs for the production of atypical polyketide extender units from enoyl-CoAs

To first identify suitable ECRs for the preparative-scale synthesis of alkyl-malonyl-CoA ester, we tested six different homologs with respect to their biosynthetic potential. Our screen included ECRs that either had been described to show relaxed substrate specificity, were part of a cluster described to produce a polyketide with atypical extender units, or contained an active site motif indicating promiscuity (Table S1). In addition, we tested an ECR variant of *C. crescentus* ($CcrC_{PAG}$) that was recently engineered from a specific into a promiscuous enzyme through three active site mutations (C146P, I169A, and F373G) [7]. The six ECRs were tested on five different enoyl-CoA thioesters of varying chain length and branching patterns; crotonyl-, hexenoyl-, 5-methylhexenoyl-, octenoyl-, and cinnamoyl-CoA. We determined the carboxylation efficiency for each enzyme with each substrate by quantifying the percentage yield of carboxylated product (compared with total product formed, including reduced side product; Figure 1A). We also measured the kinetics for each enzyme with each substrate by following the consumption of nicotinamide adenine dinucleotide phosphate (NADPH) (Figures 1B and S1; Table S2). Based on these parameters, we identified $CcrC_{PAG}$ as the most suited ECR for our biocatalytic synthesis strategy, because the enzyme showed the highest carboxylation efficiencies, as well as favorable catalytic activities for all substrates tested, with the exception of cinnamoyl-CoA. With cinnamoyl-CoA, $CcrC_{PAG}$ showed a reduced carboxylation efficiency (57%), as well as a reduced catalytic activity (<0.15 s^{-1}). Cinnamoyl-CoA also posed a challenge for other ECRs. While CinF showed higher carboxylation efficiencies (74%) compared with $CcrC_{PAG}$, the enzyme acted even slower on this substrate (k_{cat} < 0.11 s^{-1}). AntE and Ndas_0488, on the other hand, displayed 2-fold higher turnover rates with cinnamoyl-CoA compared with $CcrC_{PAG}$, albeit at reduced carboxylation efficiency.

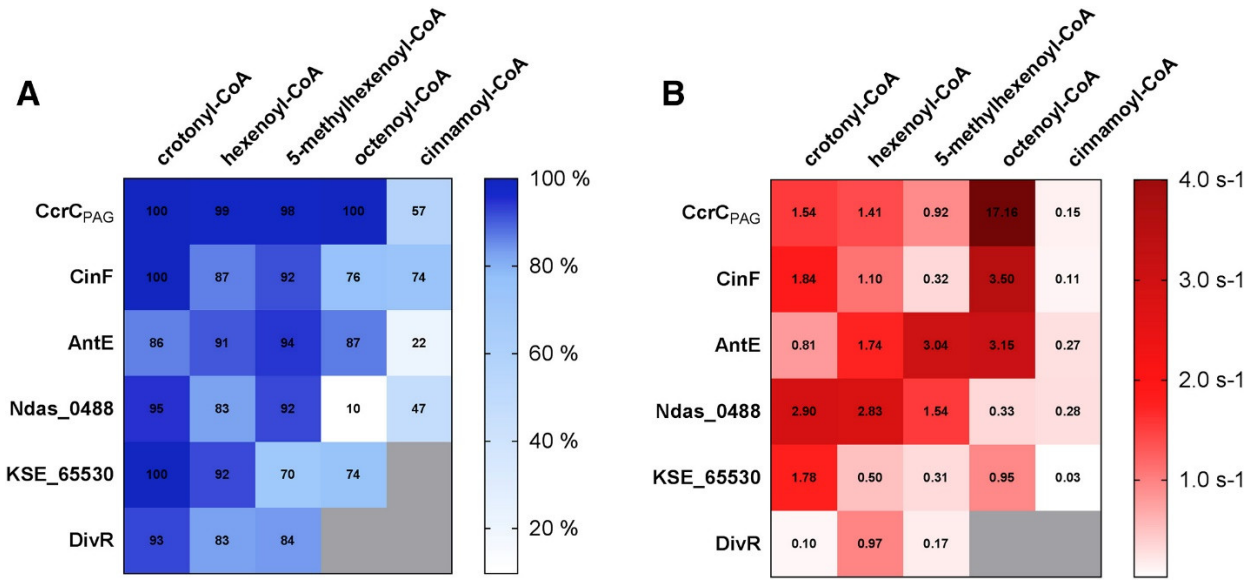


Figure 1. Characterization of ECR variants for their substrate promiscuity. (A) Percentage of carboxylation product of promiscuous ECRs dependent on the substrate compared with the reduction side reaction. Assays contained 100 mM Tris-HCl (pH 8), 100 mM $KHCO_3$, 10 µg mL^{-1} carbonic anhydrase, 150 µM NADPH, 75 µM substrate, and 1.5 µg of the respective ECR, quenched with 5% formic acid and analyzed by HPLC-MS. (B) k_{cat} of ECRs for all the substrates tested.

Biocatalytic proofreading improved atypical extender unit biosynthesis

The low carboxylation yields and reduced catalytic activities of ECRs with cinnamoyl-CoA posed a problem for the preparative-scale production of benzylmalonyl-CoA. Due to the pronounced reduction side reaction of ECRs with cinnamoyl-CoA, most of the substrate is converted into phenylpropinoyl-CoA, which represents a “dead-end” product. Thus, we sought to employ the concept of biocatalytic proofreading to our biosynthetic strategy [19]. We reasoned that an efficient regeneration of phenylpropionyl-CoA back into the substrate cinnamoyl-CoA should increase the yield of the carboxylation reaction in the synthesis assay. Therefore we tested whether the recently described acyl-CoA oxidase Acx4 from Arabidopsis thaliana [20] is able to oxidize phenylpropinoyl-CoA into cinnamoyl-CoA. A detailed kinetic characterization of Acx4 showed that the enzyme is not only active with phenlypropionyl-CoA, but accepts many different acyl-CoAs, including the ones that are relevant to this study (butyanoyl-, hexanoyl, octanoyl-, and 5-methylhexanoyl-CoA; Table 1). In the following, we tested the effect of Acx4 on the AntE-dependent carboxylation of cinnamoyl-CoA (Figure 2A). While the control reaction without

proofreading enzyme produced only 19% of benzylmalonyl-CoA (Figure 2B), the assay containing Acx4 yielded 73% benzylmalonyl-CoA (Figure 2C). The addition of catalase to remove the potentially harmful H_2O_2 produced by Acx4 did not alter the yield of the reaction and was therefore not used in further experiments. These results demonstrated that Acx4 can be directly added as proofreading enzyme to our biosynthetic assays, and that biocatalytic proofreading by Acx4 increases carboxylation product yields by almost a factor of 4. In the following we used this route to produce benzylmalonyl-CoA at the preparative scale and high purity (Figure S3A).

Table 1. Kinetic characterization of the acyl-CoA oxidase Acx4

Substrate	k_{cat} (s^{-1})	K_M (µM)	k_{cat}/K_M ($s^{-1}M^{-1}$)
Butyl-CoA	30 ± 1.1	66 ± 9	$(4.5 \pm 0.8)\cdot10^5$
Hexanoyl-CoA	66 ± 2.1	31 ± 3	$(2.2 \pm 0.3)\cdot10^6$
Octanoyl-CoA	2.8 ± 0.3	33 ± 5	$(8.5 \pm 1.3)\cdot10^4$
5-Methylhexanoyl-CoA	7.2 ± 0.3	32 ± 7	$(2.3 \pm 0.5)\cdot10^5$
Phenylpropionyl-CoA	0.21 ± 0.01	12 ± 2	$(1.8 \pm 0.3)\cdot10^4$

Assays for butyl-CoA, hexanoyl-CoA, octanoyl-CoA, and 5-methylhexanoyl-CoA were measured at 290 nm following the double bond formation using $\Delta\varepsilon290 = 2.26\ cm^{-1}\ mM^{-1}$. The assay for phenylpropionyl-CoA was measured at 308 nm using $\Delta\varepsilon290 = 16.6\ cm^{-1}\ mM^{-1}$ as determined from the spectra of purified phenylpropionyl-CoA and cinnamoyl-CoA (Figure S2).

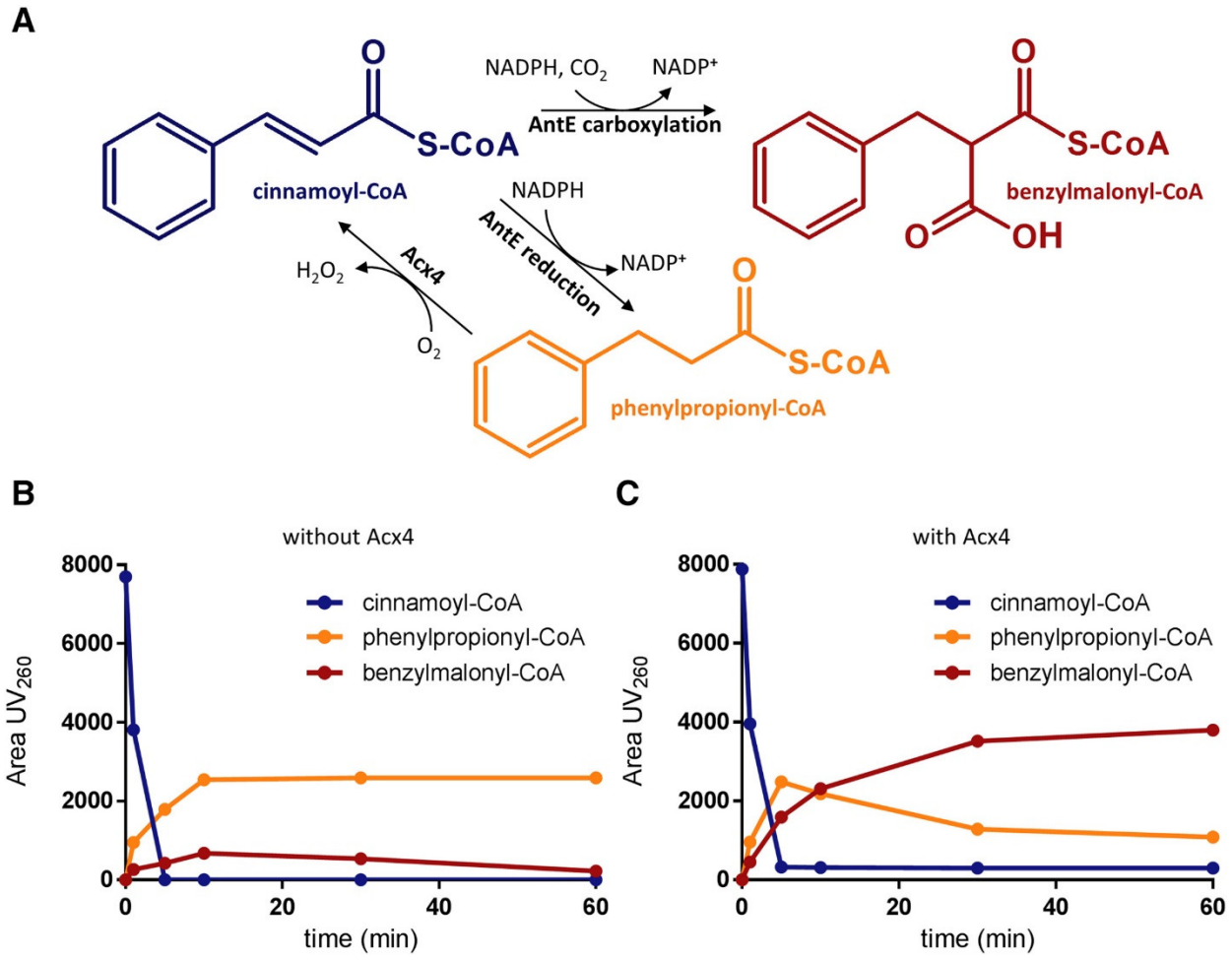


Figure 2. Cinnamoyl-CoA carboxylation using the oxidase Acx4 as a proofreading enzyme to recycle the side product dihydrocinnamoyl-CoA allows for efficient benzylmalonyl-CoA biosynthesis. (A) Scheme of the benzylmalonyl-CoA synthesis reaction containing the proofreading oxidase Acx4. (B and C) HPLC analysis of an assay containing 100 mM $NaPO_4$ buffer (pH 8), 100 mM $KHCO_3$, 100 µM cinnamoyl-CoA, 2 mM NADPH and 5.3 µM AntE. (B) Without proofreading the assay produces mainly the reduced side product dihydrocinnamoyl-CoA. HPLC assay analog to B containing additionally 2.6 µM of the proofreading enzyme Acx4.(C) Using the proofreading oxidase the assay now mainly yields benzylmalonyl-CoA.

A convenient one-pot, preparative-scale biosynthesis of atypical extender units from acyl-CoAs

Acx4 was not only useful for biocatalytic proofreading, but also opened an alternative route to the synthesis of atypical extender units. We noticed that, when promiscuous Acx4 and promiscuous $CcrC_{PAG}$ are combined, it becomes possible to synthesize alkyl-malonyl-CoAs directly from the corresponding saturated acyl-CoAs. Acyl-CoAs are

chemically easily accessible from free CoA and the respective saturated acids [21, 22], which are commercially more available and cheaper than their desaturated counterparts. We demonstrated this chemo-biocatalytic route in the following. First, we chemically coupled free CoA with butanoic-, hexanoic, 5-methylhexanoic-, and phenylpropionic acid in preparative scale [22]. Then, the individual synthesis assays were lyophilized, resolved in buffer (100 mM Tris-HCl (pH 7.5), 100 mM $KHCO_3$) to approximately 1 mM acyl-CoA, and incubated with Acx4 (2.4 µM), $CcrC_{PAG}$ (0.8 µM) as well as NADPH (10 mM) for 120 min at 30°C (see Figure S2). Progress of the individual reactions was followed by high-performance liquid chromatography-mass spectrometry (HPLC-MS) (Figure 3), assays were quenched after 120 min, CoA esters were purified via HPLC, lyophilized, and yields were determined (Table S3). Butylmalonyl-, hexylmalonyl-, and 3-methylbutylmalonyl-CoA were produced at yields of >95% from the corresponding acyl-CoA ester, and around 30% from the used CoA after HPLC purification. Only the biosynthesis of benzylmalonyl-CoA did not run to full completion under the chosen conditions, even when the concentration of Acx4 and $CcrC_{PAG}$ was increased (Figure 3E). Nevertheless, we could use this chemo-biosynthetic route to prepare three alkyl-malonyl-CoA extender units in the mg scale and at purity >95%, as judged by HPLC-MS (Figure S4; Table 1).

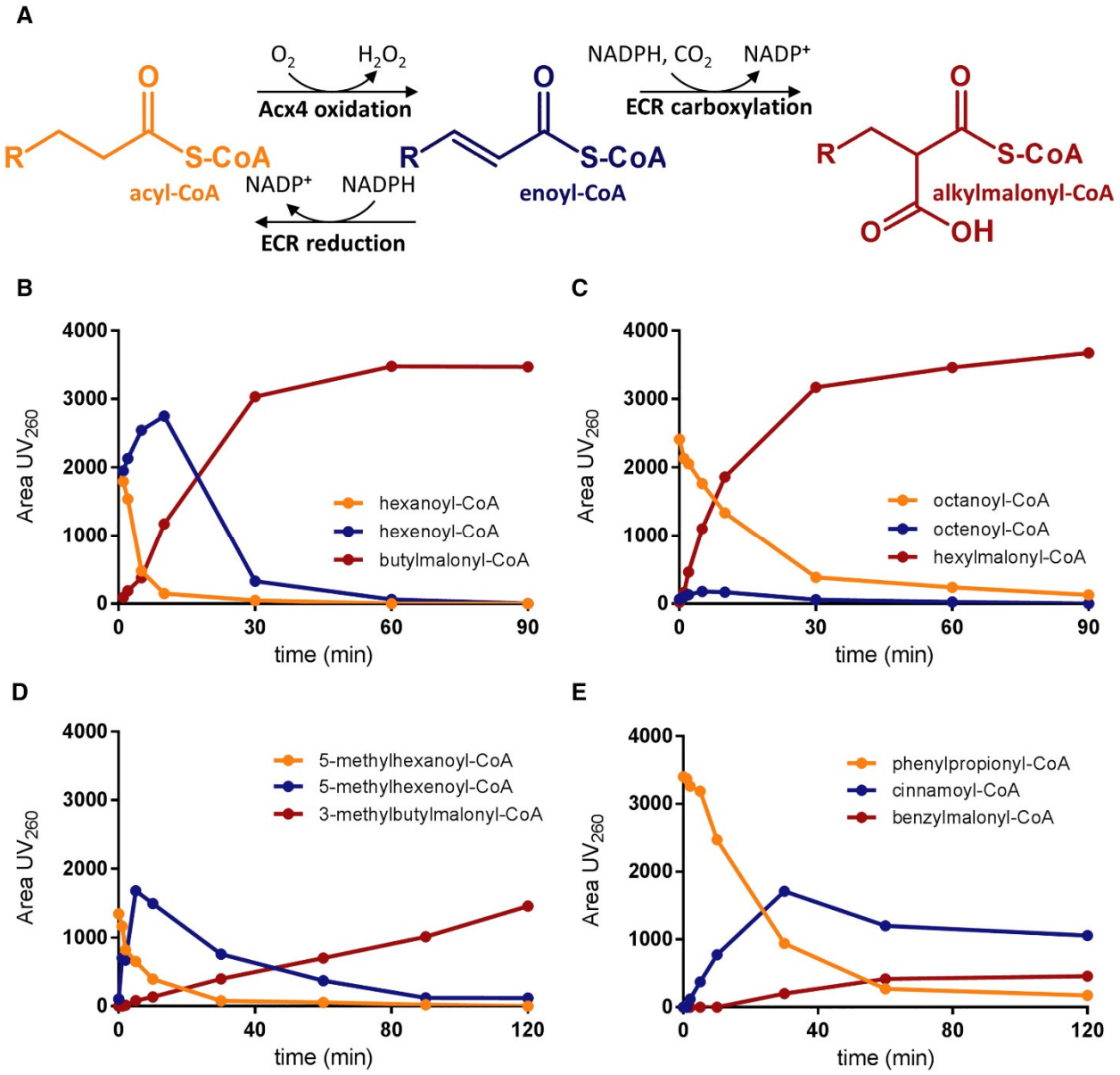


Figure 3. One-pot synthesis of atypical polyketide extender units using the acyl-CoA oxidase Acx4 and the reductive carboxylase CcrC$_{PAG}$. (A) General reaction scheme. Assays contained 100 mM Tris-HCl (pH 7.5), 100 mM $KHCO_3$, 10 mM NADPH, approximately 1 mM acyl-CoA from the CDI reaction mixture, 2.4 µM Acx4, and 0.8 µM CcrC$_{PAG}$. (B) HPLC analysis of butylmalonyl-CoA synthesis. (C) HPLC analysis of hexylmalonyl-CoA synthesis. (D) HPLC analysis of 3-methylbutylmalonyl-CoA synthesis. (E) HPLC analysis of benzylmalonyl-CoA synthesis using 5-fold increased CcrC$_{PAG}$ and 8-fold increased Acx4 concentrations compared with the other assays.

Competition assays verify selectivity-conferring active site residues of PKSs

The incorporation behavior of PKSs in the presence of a large variety of extender units at the same time has been investigated only sparsely so far. This prompted us to utilize our set of six diverse alkyl-malonyl-CoAs in PKS competition assays. As model system, we used a modified DEBS PKS *in vitro* model system that produces triketides [18]. We first tested substrate specificity of the wild-type (WT) modules of the model system in a competition experiment containing 1 mM of all extender units; the natural substrate methylmalonyl-CoA, as well as the unnatural extender units malonyl-, ethylmalonyl-, butylmalonyl-, 3-methylbutylmalonyl-, and hexylmalonyl-CoA. Samples were quenched after 60 min incubation at 30°C, and triketides were analyzed by HPLC-electrospray ionization-TOF. While the WT modules showed incorporation of ethyl- and butylmalonyl-CoA, they did not accept 3-methylbutyl-, hexyl-, or malonyl-CoA (Figure 4A), indicating that slightly larger extender units are able to enter the WT modules of the triketide model system when provided at equal concentrations as the natural extender methylmalonyl-CoA. This is well in line with earlier reports on the promiscuity of DEBS WT modules tested with individual atypical extender units [23].

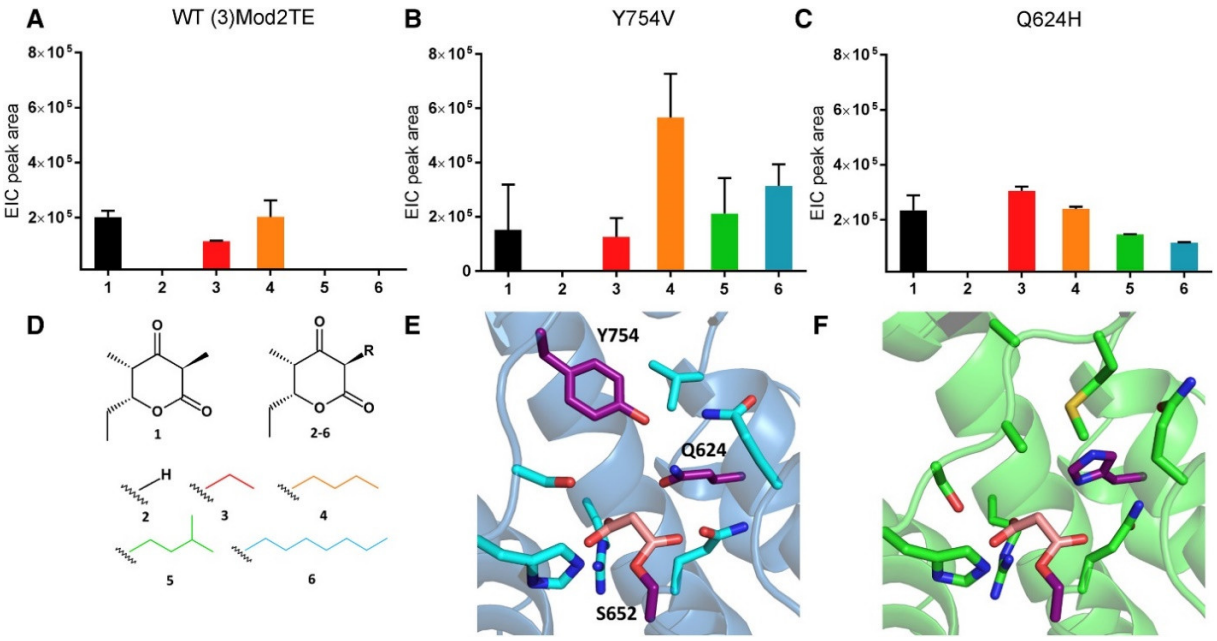


Figure 4. *In vitro* extender unit promiscuity screen for triketide biosynthesis of the DEBS polyketide synthase. The WT systems production was compared with two variants that were altered in the AT domain of module 2. Assays contained 1 mM each of the natural extender unit methylmalonyl-CoA (1) and of the

atypical malonyl-CoA (2), ethylmalonyl-CoA (3), butylmalonyl-CoA (4), 3-methylbutylmalonyl-CoA (5), and hexylmalonyl-CoA (6). Assays were incubated for 60 min and quenched with 5% formic acid. (A) Triketides produced by the WT modules of the *in vitro* DEBS triketide system. (B) Triketides produced by the Y754V variant of the AT of module 2 of the DEBS triketide system. (C) Triketides produced by the Q624H variant of the AT of module 2. Both variants additionally incorporate 3-methylbutylmalonyl-CoA (5) and hexylmalonyl-CoA (6). (D) Scheme of the produced triketides. (E) Model of the AT active site of DEBS module 2 generated with SWISS-MODEL[24] with the [KS3][AT3] didomain crystal structure (2QO3) as template[25]. Highlighted in purple are serine 652, which carries the malonyl-CoA extender unit, as well as tyrosine 754 and glutamine 624, which line the alkyl-malonyl-CoA binding pocket and which were targeted by site-directed mutagenesis in this study. (F) Model of the promiscuous AT4 from the reveromycin PKS using the same template. Highlighted is the conserved serine that carries the malonyl-CoA extender unit, as well as the histidine that replaces the conserved glutamine of methyl-CoA- and ethylmalonyl-CoA-specific ATs (i.e., Q624 of DEBS AT2). The error bars represent standard deviation of two independent assays.

Next, we tested competitive incorporation of atypical extender units into a triketide model system with altered selectivity of the AT domain of module 2. To that end we targeted Y754, which is located in the YASHS motif of the AT domain, and which was previously described to confer methylmalonyl-CoA selectivity [26]. We substituted Y754 with valine to increase selectivity of the AT toward longer side chains [27]. The Y745V showed increased incorporation of atypical extender units in the competitive assay, again in line with earlier findings (Figure 4B). Having demonstrated that our multi-substrate competition assay can be used to screen extender unit selectivity, we decided to target a, to our knowledge, sofar uncharacterized glutamine (Q624 in DEBS module 2), which is highly conserved in methyl-CoA- and ethylmalonyl-CoA-incorporating ATs. In AT4 of the reveromycin PKS [28], which displays relaxed selectivity toward longer and branched alkyl-malonyl-CoA extender units, this canonical glutamine is replaced with histidine. The Q624H variant showed increased incorporation of atypical extender units in the competitive assay, indicating the importance of this residue for substrate selectivity (Figures 4C and 4F).

Taken together our results demonstrate the potential of the developed synthesis route to prepare a set of atypical alkyl-malonyl-CoA extender units that can be used for *in vitro* screening and site-directed mutagenesis of PKSs to assess their substrate selectivity. Our chemo-biosynthetic route can be easily extended for the synthesis of a

large library of atypical extender units. Along with recently established *in vitro* polyketide systems, this will open the way to characterize large libraries of mutants in the AT domain, eventually leading to a detailed understanding of the function of active site residues that contribute to substrate selectivity in PKSs.

Acknowledgments

The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported under contract no. DE-AC02-05CH11231. This work was funded by the Max Planck Society. This project was further supported by the LOEWE program (Landes-Offensive zur Entwicklung wissenschaftlich-ökonomischer Exzellenz) of the state of Hessen and was conducted within the framework of the MegaSyn Research Cluster.

References

1. Chan, Y. A.; Podevels, A. M.; Kevany, B. M.; Thomas, M. G., Biosynthesis of polyketide synthase extender units. *Natural product reports* **2009,** *26* (1), 90-114.
2. Ray, L.; Moore, B. S., Recent advances in the biosynthesis of unusual polyketide synthase substrates. *Natural product reports* **2016,** *33* (2), 150-161.
3. Moore, B. S.; Hertweck, C., Biosynthesis and attachment of novel bacterial polyketide synthase starter units. *Natural product reports* **2002,** *19* (1), 70-99.
4. Erb, T. J.; Berg, I. A.; Brecht, V.; Müller, M.; Fuchs, G.; Alber, B. E., Synthesis of C5-dicarboxylic acids from C2-units involving crotonyl-CoA carboxylase/reductase: the ethylmalonyl-CoA pathway. *Proceedings of the National Academy of Sciences* **2007,** *104* (25), 10631-10636.
5. Erb, T. J.; Brecht, V.; Fuchs, G.; Müller, M.; Alber, B. E., Carboxylation mechanism and stereochemistry of crotonyl-CoA carboxylase/reductase, a carboxylating enoyl-thioester reductase. *Proceedings of the National Academy of Sciences* **2009,** *106* (22), 8871-8876.
6. Wilson, M. C.; Moore, B. S., Beyond ethylmalonyl-CoA: the functional role of crotonyl-CoA carboxylase/reductase homologs in expanding polyketide diversity. *Natural product reports* **2012,** *29* (1), 72-86.
7. Peter, D. M.; Schada von Borzyskowski, L.; Kiefer, P.; Christen, P.; Vorholt, J. A.; Erb, T. J., Screening and engineering the synthetic potential of carboxylating reductases from central metabolism and polyketide biosynthesis. *Angewandte Chemie International Edition* **2015,** *54* (45), 13457-13461.
8. Liou, G. F.; Khosla, C., Building-block selectivity of polyketide synthases. *Current opinion in chemical biology* **2003,** *7* (2), 279-284.

9. Yan, J.; Hazzard, C.; Bonnett, S. A.; Reynolds, K. A., Functional modular dissection of DEBS1-TE changes triketide lactone ratios and provides insight into Acyl group loading, hydrolysis, and ACP transfer. *Biochemistry* **2012,** *51* (46), 9333-9341.
10. Koryakina, I.; McArthur, J.; Randall, S.; Draelos, M. M.; Musiol, E. M.; Muddiman, D. C.; Weber, T.; Williams, G. J., Poly specific trans-acyltransferase machinery revealed via engineered acyl-CoA synthetases. *ACS chemical biology* **2013,** *8* (1), 200-8.
11. Koryakina, I.; McArthur, J. B.; Draelos, M. M.; Williams, G. J., Promiscuity of a modular polyketide synthase towards natural and non-natural extender units. *Org Biomol Chem* **2013,** *11* (27), 4449-58.
12. Tang, Y.; Tsai, S.-C.; Khosla, C., Polyketide chain length control by chain length factor. *Journal of the American Chemical Society* **2003,** *125* (42), 12708-12709.
13. Koryakina, I.; Kasey, C.; McArthur, J. B.; Lowell, A. N.; Chemler, J. A.; Li, S.; Hansen, D. A., Inversion of Extender Unit Selectivity in the Erythromycin Polyketide Synthase by Acyltransferase Domain Engineering. **2017,** *12* (1), 114-123.
14. Yuzawa, S.; Deng, K.; Wang, G.; Baidoo, E. E.; Northen, T. R.; Adams, P. D.; Katz, L.; Keasling, J. D., Comprehensive in vitro analysis of acyltransferase domain exchanges in modular polyketide synthases and its application for short-chain ketone production. *ACS synthetic biology* **2017,** *6* (1), 139-147.
15. Koryakina, I.; Williams, G. J., Mutant malonyl-CoA synthetases with altered specificity for polyketide synthase extender unit generation. *Chembiochem : a European journal of chemical biology* **2011,** *12* (15), 2289-93.
16. Pohl, N. L.; Hans, M.; Lee, H. Y.; Kim, Y. S.; Cane, D. E.; Khosla, C., Remarkably broad substrate tolerance of malonyl-CoA synthetase, an enzyme capable of intracellular synthesis of polyketide precursors. *J Am Chem Soc* **2001,** *123* (24), 5822-3.
17. Zhang, L.; Mori, T.; Zheng, Q.; Awakawa, T.; Yan, Y.; Liu, W.; Abe, I., Rational control of polyketide extender units by structure-based engineering of a crotonyl-CoA carboxylase/reductase in antimycin biosynthesis. *Angewandte Chemie International Edition* **2015,** *54* (45), 13462-13465.
18. Lowry, B.; Robbins, T.; Weng, C. H.; O'Brien, R. V.; Cane, D. E.; Khosla, C., In vitro reconstitution and analysis of the 6-deoxyerythronolide B synthase. *J Am Chem Soc* **2013,** *135* (45), 16809-12.
19. Sun, J.; Jeffryes, J. G.; Henry, C. S.; Bruner, S. D.; Hanson, A. D., Metabolite damage and repair in metabolic engineering design. *Metabolic engineering* **2017,** *44*, 150-159.
20. Schwander, T.; von Borzyskowski, L. S.; Burgener, S.; Cortina, N. S.; Erb, T. J., A synthetic pathway for the fixation of carbon dioxide in vitro. *Science* **2016,** *354* (6314), 900-904.
21. Agarwal, V.; Diethelm, S.; Ray, L.; Garg, N.; Awakawa, T.; Dorrestein, P. C.; Moore, B. S., Chemoenzymatic synthesis of acyl coenzyme A substrates enables in situ labeling of small molecules and proteins. *Organic letters* **2015,** *17* (18), 4452-4455.
22. Peter, D. M.; Vögeli, B.; Cortina, N. S.; Erb, T. J., A chemo-enzymatic road map to the synthesis of CoA esters. *Molecules* **2016,** *21* (4), 517.
23. Dunn, B. J.; Watts, K. R.; Robbins, T.; Cane, D. E.; Khosla, C., Comparative analysis of the substrate specificity of trans- versus cis-acyltransferases of assembly line polyketide synthases. *Biochemistry* **2014,** *53* (23), 3796-806.

24. Bienert, S.; Waterhouse, A.; de Beer, T. A.; Tauriello, G.; Studer, G.; Bordoli, L.; Schwede, T., The SWISS-MODEL Repository—new features and functionality. *Nucleic acids research* **2017,** *45* (D1), D313-D319.
25. Tang, Y.; Chen, A. Y.; Kim, C.-Y.; Cane, D. E.; Khosla, C., Structural and mechanistic analysis of protein interactions in module 3 of the 6-deoxyerythronolide B synthase. *Chemistry & biology* **2007,** *14* (8), 931-943.
26. Reeves, C. D.; Murli, S.; Ashley, G. W.; Piagentini, M.; Hutchinson, C. R.; McDaniel, R., Alteration of the substrate specificity of a modular polyketide synthase acyltransferase domain through site-specific mutations. *Biochemistry* **2001,** *40* (51), 15464-70.
27. Smith, L.; Hong, H.; Spencer, J. B.; Leadlay, P. F., Analysis of specific mutants in the lasalocid gene cluster: evidence for enzymatic catalysis of a disfavoured polyether ring closure. *Chembiochem : a European journal of chemical biology* **2008,** *9* (18), 2967-2975.
28. Takahashi, S.; Toyoda, A.; Sekiyama, Y.; Takagi, H.; Nogawa, T.; Uramoto, M.; Suzuki, R.; Koshino, H.; Kumano, T.; Panthee, S., Reveromycin A biosynthesis uses RevG and RevJ for stereospecific spiroacetal formation. *Nature chemical biology* **2011,** *7* (7), 461.

Supplemental Information

Table S1 (related to Figure 1): Overview of the selected promiscuous ECRs

Enzyme	Phylum	Family	Organism	Promiscuity motif			Accession code	Source
$CcrC_{PAG}$	Proteo-bacteria	Caulobacter-iaceae	Caulobacter crescentus	P	A	G	WP_010920921	[1]
CinF	Actino-bacteria	Streptomyceta-ceae	Streptomyces sp. JS360	P	A	G	WP_067247071.1	[2]
AntE	Actino-bacteria	Streptomyceta-ceae	Streptomyces sp. NRRL2288	A	V	V	AGG37751.1	DOE-JGI (this work)
Ndas_0488	Actino-bacteria	Nocardiopsa-ceae	Nocardiopsis dassonvillei	P	V	F	WP_013151542.1	DOE-JGI (this work)
KSE_65530	Actino-bacteria	Streptomyceta-ceae	Kitasatospora setae	T	A	G	YP_004908268.1	DOE-JGI (this work)
DivR	Actino-bacteria	Streptomyceta-ceae	Streptomyces sp. HKI0576	A	A	V	CCA30183	DOE-JGI (this work)

Table S2 (related to Figure 1): k_{cat} and K_M of the ECR variants towards the tested substrates. Values with yellow background were fitted using substrate inhibition. ND = not determinable, errors are given as standard error.

k_{cat} [s^{-1}]					
	crotonyl-CoA	hexenoyl-CoA	5-methylhexenoyl-CoA	octenoyl-CoA	cinnamoyl-CoA
$CcrC_{PAG}$	1.5 ± 0.1	1.4 ± 0.1	0.9 ± 0.1	17 ± 1	0.15 ± 0.01
CinF	1.84 ± 0.03	1.10 ± 0.04	0.32 ± 0.03	3.5 ± 0.7	0.113 ± 0.003
AntE	0.81 ± 0.04	1.7 ± 0.1	3.0 ± 0.1	3.1 ± 0.2	0.27 ± 0.01
Ndas_0488	2.9 ± 0.5	2.8 ± 0.5	1.5 ± 0.1	0.33 ± 0.01	0.28 ± 0.02
KSE_65530	1.78 ± 0.05	0.50 ± 0.02	0.6 ± 0.1	1.0 ± 0.4	0.03 ± 0.003
DivR	0.089 ± 0.003	1.0 ± 0.2	0.17 ± 0.01	ND	ND
K_M [µM]					
	crotonyl-CoA	hexenoyl-CoA	5-methylhexenoyl-CoA	octenoyl-CoA	cinnamoyl-CoA
$CcrC_{PAG}$	525 ± 85	74 ± 10	153 ± 45	75 ± 15	3 ± 2
CinF	154 ± 10	16 ± 2	53 ± 11	11 ±4	52 ± 6
AntE	218 ± 31	103 ± 18	125 ± 18	133 ± 18	42 ± 6
Ndas_0488	38 ± 17	58 ± 17	30 ± 5	45 ± 6	160 ± 35
KSE_65530	111 ± 8	33 ± 5	36 ± 17	100 ± 60	ND
DivR	10 ± 2	96 ± 30	56 ± 9	ND	ND

Table S3 (related to Figure 3): Yields of extender unit biosynthesis

Type of acyl-CoA	Yield for chemical coupling	Yield of one pot synthesis	Overall yield after HPLC purification
butylmalonyl-CoA	57%	99%	27%
3-methylbutylmalonyl-CoA	50%	100%	36%
hexylmalonyl-CoA	72%	93%	28%
benzylmalonyl-CoA	49%	27%	-
benzylmalonyl-CoA (one pot from cinnamoyl-CoA without proofreading)	-	8%	-
benzylmalonyl-CoA (one pot from cinnamoyl-CoA)	-	73%	37%

Table S4 (related to Figure 1): Sequences of enzymes used in this study

CcrC$_{PAG}$	DNA	ATGGGCCATCATCATCATCATCATCATCATCACAGCAGCGGCCATATCGAAGGTCGTCATA TGACGATTCAGACGCTTGAGACCACCGCGCTGAAGGACCTGTACGAGATCGGTGAGATCCCGC CGGCCTTCCACGTGCCGAAAACCATGTACGCCTGGAGCATCCGCAAGGAGCGCCACGGCAAG CCGACCCAGGCCATGCAGGTCGAGGTGGTGCCCACCTGGGAGATCGGCGAGGACGAGGTGCT GGTGCTCGTGATGGCCGCCGGCGTCAATTACAACGGGGTCTGGGCCGCTCTGGGCGAGCCGA TCAGCCCGCTGGACGGCCACAAGCAGCCTTTCCACATCGCCGGGTCCGACGCCTCGGGCATC GTCTGGAAGGTCGGCGCCAAGGTGAAGCGGTGGAAGCTGGGCGACGAGGTCGTCATCCACCC GAACCAGGACGACGGCGACGACGAGGAGTGCAACGGCGGCGACCCGATGTTCTCGTCCAGCC AGCGCGCGTGGGGCTACGAGACGCCGGACGGCAGCTTCGCCCAGTTCTGCCGGGTGCAGTC GCGCCAGCTGCTGCCGCGCCCCAAGCACCTGACCTGGGAAGAGAGCGCCTGCTACACCCTGA CCCTGGCCACCGCCTACCGCATGCTGTTCGGCCACAAGCCGCATGAGCTGAAGCCCGGCCAG AACGTGCTGGTCTGGGGCGCCTCGGGCGGTCTTGGCGTCTTCGCCACCCAGCTGGCCGCCGT GGCCGGCGCCAACGCCATCGGCGTGGTGTCCTCCGAGGATAAGCGCGAGTTCGTGCTGTCGA TGGGCGCCAAGGCGGTGCTGAACCGGGGCGAGTTCAACTGCTGGGGCCAGCTGCCGAAGGTC AACGGCCCCGAGTTCAACGACTACATGAAGGAGAGCCGCAAGTTCGGCAAAGCCATCTGGCAG ATCACCGGAAACAAGGACGTCGACATGGTGTTCGAGCACCCTGGCGAGCAGACCTTCCCGGTG TCGGTGTTCCTGGTCAAGCGCGGCGGCATGGTGGTGATCTGCGCCGGCACGACGGGCTTCAA CCTGACCATGGACGCCCGCTTCCTGTGGATGCGCCAGAAGCGCGTGCAAGGGTCGCACGGCG CCAACCTGATGCAGGCCAGCGCCGCCAACCAGCTGGTCATCGACCGCCGCGTCGATCCCTGC CTGTCGGAAGTCTTCCCCTGGGACCAGATCCCGGCGGCCCACGAGAAGATGCTGGCCAACCA GCACCTGCCGGGGAACATGGCCGTGCTGGTCTGCGCCCAGCGCCCCGGCCTGCGCACCTTCG AGGAAGTGCAGGAGCTGAGCGGGGCGCCATAG
	Protein	MGHHHHHHHHHHSSGHIEGRHMTIQTLETTALKDLYEIGEIPPAFHVPKTMYAWSIRKERHGKPTQA MQVEVVPTWEIGEDEVLVLVMAAGVNYNGVWAALGEPISPLDGHKQPFHIAGSDASGIVWKVGAKV KRWKLGDEVVIHPNQDDGDDEECNGGDPMFSSSQRAWGYETPDGSFAQFCRVQSRQLLPRPKHL TWEESACYTLTLATAYRMLFGHKPHELKPGQNVLVWGASGGLGVFATQLAAVAGANAIGVVSSEDK REFVLSMGAKAVLNRGEFNCWGQLPKVNGPEFNDYMKESRKFGKAIWQITGNKDVDMVFEHPGE QTFPVSVFLVKRGGMVVICAGTTGFNLTMDARFLWMRQKRVQGSHGANLMQASAANQLVIDRRVD PCLSEVFPWDQIPAAHEKMLANQHLPGNMAVLVCAQRPGLRTFEEVQELSGAP

AntE	DNA	ATGGGCCATCATCATCATCATCATCATCATCACAGCAGCGGCCATATCGAAGGTCGTCATA TGACTAAAGACCTGTACGAACTGGGCGACGCTCCGCCACTGGGTACTGCGCCGAAACAGATGT ATGCTTCCCTTATCCGCCAGGAGCGTTATGGCCGTCCGGTTGACGCTTTCCGTACCGAAGTAGT GGATGTACCGCCGGTTGGTCCGGGTCAGGTACTGGTTAAAGTAATGGCAGCTGGCGTAAACTA CAACAACGTGTGGGCGGCACTGGGTGAGCCACTGGACGTAATTGCAGCGCGCCAGAAACAGG GCGCTACCGAAGACTTTCATATCGGCGGTTCAGACCTGTCTGGCATTGTATGGGCGGTTGGTG ACGGGGTACGTCTGAAACCGGGTGCAGAGGTTGTGGTTCTGGCTTGTCGTTGGGACGAATCTG CACAGGACATCCGCCTGGGCGCTGACCCGGTTACCTCTTCTACTCAGCGTGTATGGGGCTATG AAGAAAACTACGGTTCTTTCGCTCAGTTTGCAGTTGTAGACGAATATATGTGCCACCCGAAACC GCAGCGCCTGAGCTGGGCTGCGGCTTCTTGCTACATGGCCACCGCCGCGACCGCCTATCGTC AGCTGTTCGGCTGGGAACCGCACACTGTTCGTCCGGGTGACCCGGTTCTGATCTGGGGCGGT GCAGGTGGTCTGGGCTCCATCGCGATCCAGCTGGTACGTCACGTTGGTGGCATCCCGGTAGCT GTGGTGAGCTCCGAAGAACGTGGTGAATTCTGTATGCGCCTGGGCGCGAAAGGTTGGATCGAC CGTCGTGAATTCGACCACTGGGGTCGTCTGCCGGACACCACTGACGAGGAAGCTATGAAACAG TGGCTGGACGGCGCGCGCGCTTTCGGTCGTCGTTTCTGGGAAGTTCTGGGTGAACGCCGTGC GCCGCGTATCGTTCTGGAACACTCTGGTGCTGACACCATCCCGACTTCCATCTATATGGCTGAC AACGCCGGTATGGTGGTTATCTGCGGTGGCACGACTGGCTATAACGGTGACGTTGACCTGCGT TTCCTGTGGATGCGTCAAAAGCGCCTGCAGGGTAGTCACGTAGCTTCTGCACGCGAAGCACGT GAAATCACTCGCCTGATCGACCAAGGTGCAATCGACCCGTGCCTGTCCCGTACCTTCGGTTTC GAAGAGATCGGCCTGGCTCACCAGCTGATCCACGATAACCAGCACCCGTCTGGTAACATGGCT GTTCTGGTTAACGCAACGGAGTAA
	Protein	MGHHHHHHHHHHSSGHIEGRHMTKDLYELGDAPPLGTAPKQMYASLIRQERYGRPVDAFRTEVVD VPPVGPGQVLVKVMAAGVNYNNVWAALGEPLDVIAARQKQGATEDFHIGGSDLSGIVWAVGDGVR LKPGAEVVVLACRWDESAQDIRLGADPVTSSTQRVWGYEENYGSFAQFAVVDEYMCHPKPQRLS WAAASCYMATAATAYRQLFGWEPHTVRPGDPVLIWGGAGGLGSIAIQLVRHVGGIPVAVVSSEERG EFCMRLGAKGWIDRREFDHWGRLPDTTDEEAMKQWLDGARAFGRRFWEVLGERRAPRIVLEHSG ADTIPTSIYMADNAGMVVICGGTTGYNGDVDLRFLWMRQKRLQGSHVASAREAREITRLIDQGAIDP CLSRTFGFEEIGLAHQLIHDNQHPSGNMAVLVNATE
Ndas_0488	DNA	ATGGGCCATCATCATCATCATCATCATCATCACAGCAGCGGCCATATCGAAGGTCGTCATA TGCCGAAAGACCTGTACGACCTCGGCGAAGTTCCGCCGCTGGGTCACGTTCCGGCACGTATGC ATGCTTTCACTCTGCGTCGCGAACGTTACGGTCAGCCGCGTGAAGCCTTCGCTCGTGAGGTAG TTCCGGTCCCGCGTCTGGAACCGGGTCACGTACTGGTTTACACCATGAGCGCGGGCATTAACT ATAACAACGTATGGGCTGCAATGGGTCACCCGGTTGACGTAATCGCAATGCGTGAAAAGGTGG GTGCGGCTGAAGAGTACCACATCGGCGGTTCTGACGGTGCCGGTGTTGTGTGGGCTGTTGGT GAAGGTGTGCGCGGTGTTGAGGTTGGTGACCACGTAATCATCGCACCGGGCCAGTGGGATGA ATCCGCTGACGACATTCGTATGGGTCGCGACCCGATCGCATCCCAGTCTATGCGTGTTTGGGG TTACGAGAACAACTTTGGCTCCTTCGGTCAGTTCACTCTGGTACGTGACATCCAGTGTCACCGC AAACCGGAAAACCTCCCGTGGGACGTTGCAGGCGGTTTCCTGGTATCTGCTGCCACCGCTTAC CGTCAGCTGTTCGGTTGGGAACCGAACGTTGTACGCCCGGGTGACCCGGTTCTGATCTGGGGT GGTGCTGGTGGTCTGGGCACATCCGCAATCCAGCTGGCTCGTCAAGTTGGTGGCAAGCCGGTA GCTGTGGTTTCCACCGAAGACAAAGCTCGTATGTGCCGTGAACTGGGCGCAGTAGGCGTTATC CAGCGTACTGAATTCGACCATTGGGGTCGTGTGCCGGACGAGGGCGACACACAGGCTTACGCA TCCTGGATGCGTGGTGTGCGCGCATTCGGCAAGCGCTTCTGGGAAGAACTGGGCGAGCGTCG TGCACCACGTATCGTTTTCGAACACACTGGTGCAGACACTCTGCCGACTTCTCTGTACCTGTGC GACAATGCTGGCATGGTTGTACTGTGCGGTGCAACCTCTGGTTTCCAGGCAGACGTTGATCTG CGCTTCCTGTGGATGCGTCTGAAACGCCTGCAGGGTTCCCACTTCGCATCTCCGGCTCAGTGC CGCATGGTAATCGACCTGGTAGCAGGTGGCCAGCTGGATCCGTGTGTTACTCGTATCGTGGAG TTCGACGAAATTGGTGAAGCGCATCAGCTGATCCGCGACAACGCTCAGCCGCCGGGCAACATG TCTGCTCTGGTAAACGCACGTGCAGGCCAGACTGGTCTGGACGTTTAA
	Protein	MGHHHHHHHHHHSSGHIEGRHMPKDLYDLGEVPPLGHVPARMHAFTLRRERYGQPREAFAREVV PVPRLEPGHVLVYTMSAGINYNNVWAAMGHPVDVIAMREKVGAAEEYHIGGSDGAGVVWAVGEGV RGVEVGDHVIIAPGQWDESADDIRMGRDPIASQSMRVWGYENNFGSFGQFTLVRDIQCHRKPENL PWDVAGGFLVSAATAYRQLFGWEPNVVRPGDPVLIWGGAGGLGTSAIQLARQVGGKPVAVVSTED KARMCRELGAVGVIQRTEFDHWGRVPDEGDTQAYASWMRGVRAFGKRFWEELGERRAPRIVFEH TGADTLPTSLYLCDNAGMVVLCGATSGFQADVDLRFLWMRLKRLQGSHFASPAQCRMVIDLVAGG QLDPCVTRIVEFDEIGEAHQLIRDNAQPPGNMSALVNARAGQTGLDV
KSE_65530	DNA	ATGGGCCATCATCATCATCATCATCATCATCACAGCAGCGGCCATATCGAAGGTCGTCATA TGCAGATGGACGAACTGACAGAAGCGGTTCTGCGTGACGCTCCGCCGGAGGAACTGTCTCGTC TGCCGCTGCCGAAAGACTATACCGCAGCACACCTGCTGCGCCAGGACGTAGAAATGTTCCATG GCGTGGCTGATAAAGACGTTCGCCGCAGCCTGCATGTTGGTCGCGTACCGCTGCCGGAACTG

		GCTCCGGACGAAGTGCTCGTTGCTGTTATGGCGTCTTCCGTTAACTACAACACCGTATGGTCTG CAACTTTCGAGCCGGTTTCCACCTTCGACGCTCTGCGTCGTTACGCTCGCAGCGGTGGCTGGC AGGCACGCCACGATCAGCCGCACCAGGTAATCGGCTCTGACGCAGCTGGTGTTATCGTACGTA CGGGCGCTGGCGTACGCCGTTGGCAGGTAGGTGATCATGTTGCGGTATCTACCGCTGTGGTG GACGACCAGGACCCAGTTACTCACACTGACGGCATGCTGGGTGCAGATCAGAAGGCATGGGGT TACGAAACCAACTTCGGCGGCTTGGCTCACTACACTGTTGTTCGTGCGTCTCAGCTGATCGCTA AACCGCCACACCTGACCTGGGAAGAAACTGCTAGCATCCCGCTGTGCGGTGGCACAGCATACC GTATGTTAGTTTCTGAAAAAGGTGCTCGTATCAAACAGGGTGACATCGTACTTATCTGGGGTGC GTCCGGTGGTCTGGGTGCATTCGCTGTTCAGCTGGTTAAGAACGGTGGTGGTATTCCGGTAGG TGTTGTAAACTCTGAACGTAAAGCAGAACTTGTACGTCGTCTGGGCTGTGACGTAGTTATCAAC CGTGAAGAAATTGGTATCGGTAAAGCACCGGAAAGCCCGCAGGAAACTGTTGAACTGGCTAAG CGCCTGGGTCGTGCCATCCGTTCTCAGGTTGGTGAAGACCCGCACATCGTATTCGACCATGTA GGTAAAGCAACCTTTGGCATCAGCGTTATCGTTGCGCGTCGTGGTGGTACCGTTGTAACTTGCG GCTCTTCCACCGGCTACCAGCACACCTTCGATAACCGTTACTTCTGGATGAACCTGAAGCGTAT CGTAGGTTCTCATGGTATGAACCTGGGCGAGGCTGCTGAAATGATGCGTCTGTACAAACTGGG TCTGCTGGCACCGGTTGTTTCTCGCACTTACCCACTGGCAGAAGTTGGCGAAGCGGCTCGTCT GGTTCAAAACAACCAGCACACCGGCAAAATCGGTGTGCTGTGTCTGGCAGACCAACCGGGTCT GGGTGTTACCGACCCGGCAACTCGCGCTCGTCTGGGCGAGGACTGGCTGCGTCCGCTGGCAG AGGACCGTGTTCCGGCTCTGGCTGGCTAA
	Protein	MGHHHHHHHHHHSSGHIEGRHMQMDELTEAVLRDAPPEELSRLPLPKDYTAAHLLRQDVEMFHGV ADKDVRRSLHVGRVPLPELAPDEVLVAVMASSVNYNTVWSATFEPVSTFDALRRYARSGGWQARH DQPHQVIGSDAAGVIVRTGAGVRRWQVGDHVAVSTAVVDDQDPVTHTDGMLGADQKAWGYETNF GGLAHYTVVRASQLIAKPPHLTWEETASIPLCGGTAYRMLVSEKGARIKQGDIVLIWGASGGLGAFA VQLVKNGGGIPVGVVNSERKAELVRRLGCDVVINREEIGIGKAPESPQETVELAKRLGRAIRSQVGE DPHIVFDHVGKATFGISVIVARRGGTVVTCGSSTGYQHTFDNRYFWMNLKRIVGSHGMNLGEAAEM MRLYKLGLLAPVVSRTYPLAEVGEAARLVQNNQHTGKIGVLCLADQPGLGVTDPATRARLGEDWLR PLAEDRVPALAG
DivR	DNA	ATGGGCCATCATCATCATCATCATCATCATCACAGCAGCGGCCATATCGAAGGTCGTCATA TGGCGGTGGGTTCTCTGACCGATGCACCGGCAGACGCAGTTTCTGCTGAAGAAAACGTTGGTG CAGACCTGCCGGCAGAATTCCGTGCAGTACATATCTGCGCGGAAGACGAAGCAATGTTCTCCG ACCTGTCTGATAAAGACGTACGTAAATCCCTGCGCGTTGGCCCGGTTGCTATGCCAGAACTGG CTCCGGACGAAGTTCTGATCGCAGTTATGGCATCTTCTGTTAACTACAACACCGTATGGAGCGC TCAGTTCGCACCGGTGTCTCCGTTTCGCTTCCTGGAGAAATTCGGTCGTCGTGGTGGCTGGGC TGCACGTCACGACCTGCCGCACCAGGTTCTGGGTTCTGACGCTGCTGGCGTAGTAGTGCGTAC TGGTTCCGGTGTTCGCCGCTGGCGTGTAGGTGATCACGTTGTAGTTAGCGCCGCATACGTAGA CGAACAGGACCCGGGTACGCACGCAGATGGTATGATCGGTGAAGACCAGCTGGCATGGGGCT ACGAAACTAACTTCGGTGGCATGGCTGACTATGCGGTTGCTCGCGCAAGCCAGCTGATCCCGA AACCGCCGCACCTGACCTGGGAAGAAGCTGCATCTAACACTGCATGCGCATCTACCGCGTACC GTATGCTAGTAGGTGAGCGCGGTGCTCGTATGAAACAGGGTGACGTAGTTCTGGTATGGGGTG CAGCTGGCGGTCTCGGCTCCTACGGTGTTCAGCTGGTTCGTAACGGTGGTGGCGTACCGGTG GCAGTAGTGTCTTCCGCTCGCAAAGCTGAAGCAGTTCGTCGTCTGGGTTGTGATATTGTAGTTG ACCGCGCTGAGATCGGTCTGACTGACGATCCGACCGACGACCCGGATGAGGTAATTCGTATCG GTAAACGCCTGGGTGCAATCATCCGTGAACGTACCGGTCGCGACCCGGACATCGTATTCGAGC ACATCGGTCGTGCTACCTTCGGCGTATCTGTTTTCGTTGTACGTCGCGGTGGTGTAGTAGTTAC TTGTGGTAGCTCTACCGGTTACCAGCATGTATTCGATAACCGTTATCTGTGGATGAAACTAAAAC GTGTAGTAGGCTCTCACGTTGCAAACCTGCAGGAAGCGTGGGACTGCAACCGCCTGTTTGAAT TGGGTGCAATCGTACCGACTGTTAGTGCAGTATTCCCGATGGACGAGGTGGGCGAGGCTGTTC GCATCGTGCAGAACAACCGCCACATTGGCAAAATCGCGGTACTGTGTCAGGCTGACCGCCCTG GTCTGGGTGTTACCGACCCGGAACTGCGTGCTCGTGTAGGTGGCGACGATCGCCTGAACCCG CTGCGTGGGATGACTGCGGTAGGCGAAGGTGAATAA
	Protein	MGHHHHHHHHHHSSGHIEGRHMAVGSLTDAPADAVSAEENVGADLPAEFRAVHICAEDEAMFSDL SDKDVRKSLRVGPVAMPELAPDEVLIAVMASSVNYNTVWSAQFAPVSPFRFLEKFGRRGGWAARH DLPHQVLGSDAAGVVVRTGSGVRRWRVGDHVVVSAAYVDEQDPGTHADGMIGEDQLAWGYETNF GGMADYAVARASQLIPKPPHLTWEEAASNTACASTAYRMLVGERGARMKQGDVVLVWGAAGGLG SYGVQLVRNGGGVPVAVVSSARKAEAVRRLGCDIVVDRAEIGLTDDPTDDPDEVIRIGKRLGAIIRER TGRDPDIVFEHIGRATFGVSVFVVRRGGVVVTCGSSTGYQHVFDNRYLWMKLKRVVGSHVANLQE AWDCNRLFELGAIVPTVSAVFPMDEVGEAVRIVQNNRHIGKIAVLCQADRPGLGVTDPELRARVGG DDRLNPLRGMTAVGEGE
Acx4	DNA	ATGGGCCATCATCATCATCATCATCATCATCACAGCAGCGGCCATATCGAAGGTCGTCATA TGGCAGTGTTATCGTCAGCGGATCGTGCGTCTAACGAAAAGAAAGTGAAATCCAGCTATTTCGA

		TCTTCCGCCTATGGAGATGTCTGTGGCCTTTCCACAAGCAACACCAGCCAGCACTTTTCCACCG TGTACGAGCGATTACTATCACTTCAACGATCTGCTCACCCCGGAAGAACAGGCGATCCGCAAGA AAGTGCGCGAATGCATGGAAAAGGAAGTAGCTCCGATCATGACAGAGTACTGGGAGAAAGCCG AATTTCCGTTTCACATTACCCCGAAACTGGGTGCAATGGGAGTGGCTGGTGGGAGCATCAAAG GCTATGGATGCCCCGGCTTGTCGATTACCGCGAATGCTATTGCCACCGCAGAGATTGCACGCG TTGATGCCAGTTGCTCAACGTTCATTCTGGTCCATTCAAGCCTGGGTATGCTGACCATCGCTTT GTGTGGCTCTGAAGCGCAGAAAGAGAAATACTTACCGTCCTTGGCGCAACTGAACACTGTAGC GTGTTGGGCCCTGACGGAACCGGATAACGGCTCTGACGCGTCGGGTCTGGGTACCACAGCAA CCAAAGTTGAAGGCGGTTGGAAAATCAATGGTCAGAAACGCTGGATTGGCAATAGCACGTTTGC GGATCTGCTCATTATCTTTGCCCGTAATACGACGACTAACCAGATTAACGGCTTTATCGTGAAGA AGGATGCTCCAGGCCTGAAAGCGACCAAGATTCCGAACAAAATCGGTCTGCGCATGGTTCAGA ATGGCGACATTCTTCTGCAGAACGTGTTCGTTCCGGATGAGGATCGCTTGCCCGGCGTCAACT CCTTCCAAGACACCAGTAAAGTGTTAGCCGTTAGCCGTGTCATGGTTGCCTGGCAACCTATCGG AATTTCGATGGGCATCTACGACATGTGTCATCGCTATCTGAAAGAACGGAAACAGTTTGGTGCT CCTTTAGCGGCCTTCCAGCTGAATCAACAGAAACTTGTGCAGATGCTGGGGAATGTACAAGCGA TGTTTCTTATGGGCTGGCGTCTGTGCAAACTCTATGAAACGGGCCAGATGACTCCCGGACAGG CTTCGCTGGGGAAGGCATGGATTAGTTCCAAAGCCCGTGAAACCGCGTCATTAGGTCGTGAAC TGCTGGGCGGCAATGGGATTCTGGCAGACTTTCTGGTCGCCAAAGCGTTCTGCGATTTGGAAC CGATCTATACCTATGAGGGGACTTACGACATTAACACCCTCGTAACCGGTCGCGAAGTCACGG GTATTGCGAGTTTCAAACCGGCAACACGCAGCCGGCTGTAA
	Protein	MGHHHHHHHHHHSSGHIEGRHMAVLSSADRASNEKKVKSSYFDLPPMEMSVAFPQATPASTFPPC TSDYYHFNDLLTPEEQAIRKKVRECMEKEVAPIMTEYWEKAEFPFHITPKLGAMGVAGGSIKGYGCP GLSITANAIATAEIARVDASCSTFILVHSSLGMLTIALCGSEAQKEKYLPSLAQLNTVACWALTEPDNG SDASGLGTTATKVEGGWKINGQKRWIGNSTFADLLIIFARNTTTNQINGFIVKKDAPGLKATKIPNKIG LRMVQNGDILLQNVFVPDEDRLPGVNSFQDTSKVLAVSRVMVAWQPIGISMGIYDMCHRYLKERKQ FGAPLAAFQLNQQKLVQMLGNVQAMFLMGWRLCKLYETGQMTPGQASLGKAWISSKARETASLG RELLGGNGILADFLVAKAFCDLEPIYTYEGTYDINTLVTGREVTGIASFKPATRSRL

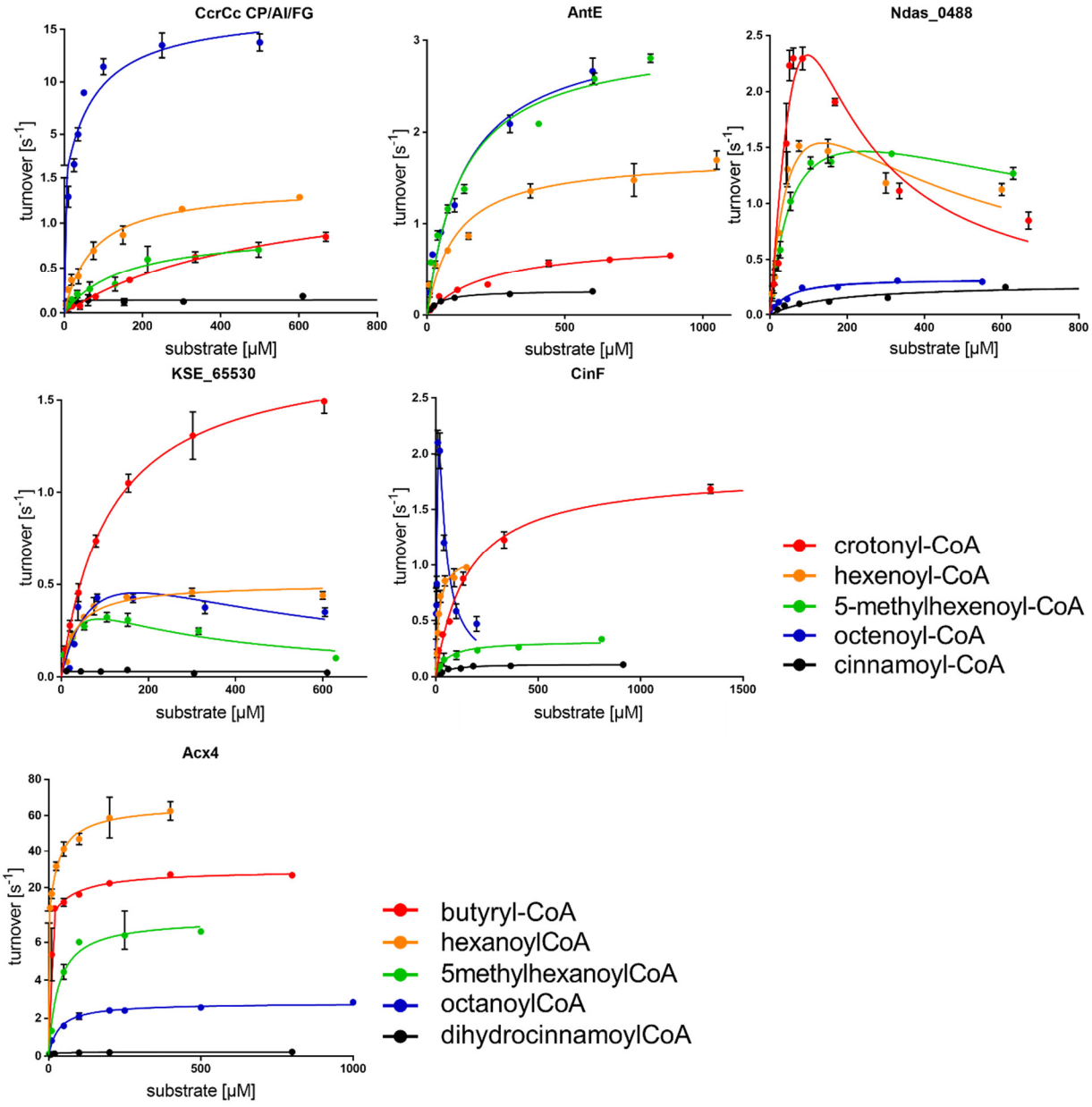


Figure S1 (related to Figure 1 and Table S2): Michaelis-Menten plots of characterized ECR variants as well as Acx4. Each point was measured as a triplicate and the error bars are given as standard deviations. Every curve was determined by a minimum of 18 points.

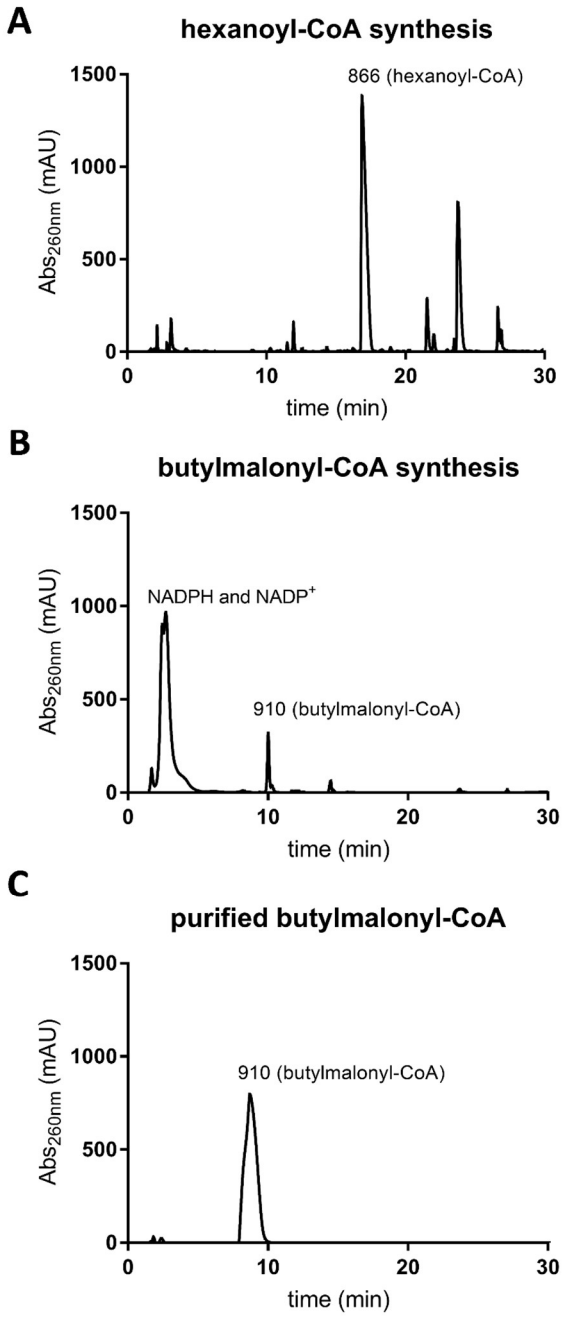


Figure S2 (related to Figure 3) HPLC traces of every step in the synthesis of butylmalonyl-CoA. (A) HPLC trace after chemical coupling. Amount of hexanoyl-CoA was assessed by comparing the absorption of hexanoyl-CoA at 260 nm to all other absorbing species in the mixture at this wavelength. (B) HPLC trace after oxidation and carboxylation of hexanoyl-CoA. Purity was assessed by the absorption of butylmalonyl-CoA compared to all other absorbing species except of the NADPH/$NADP^+$ peak at 260 nm. (C) HPLC trace of the purified butylmalonyl-CoA.

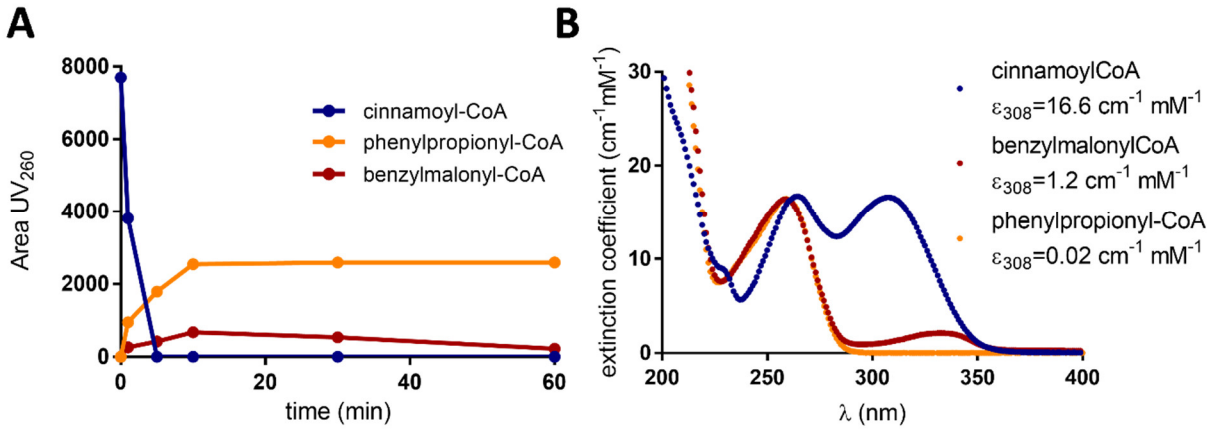
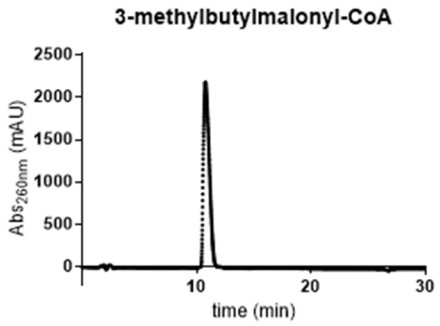
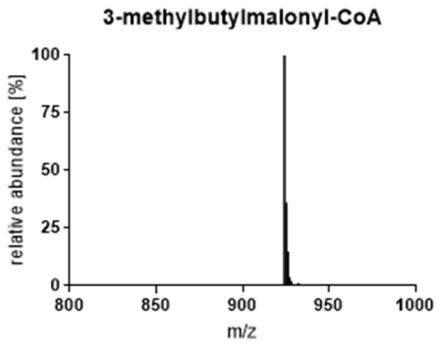
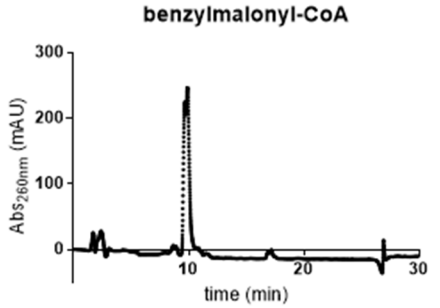
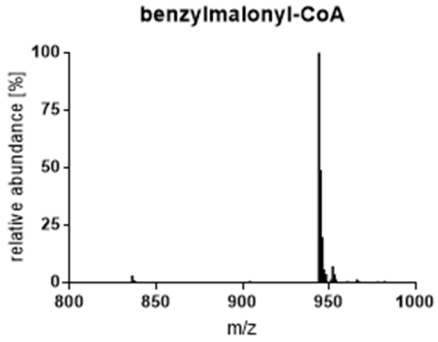


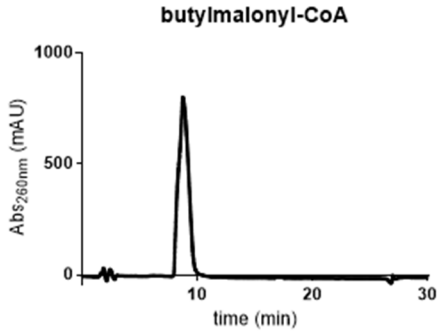
Figure S3 (related to Figure 3): Progress of up-scaled benzylmalonyl-CoA synthesis from cinnamoyl-CoA. (A) The assay contained 100 mM TrisHCl pH 8, 100 mM $KHCO_3$, 1 mM cinnamoyl-CoA, 20 mM NADPH, 40 µg mL^{-1} carbonic anhydrase, 20 µM Acx4 and 2.1 µM $CcrC_{PAG}$. (B) Extinction coefficients of cinnamoyl-CoA, dihydrocinnamoyl-CoA and benzylmalonyl-CoA.

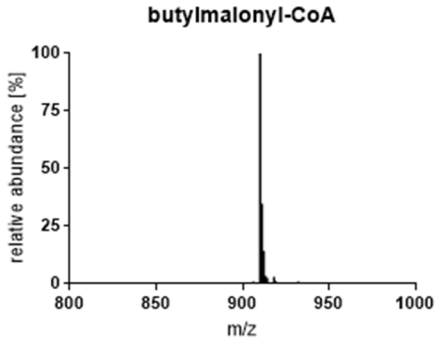

3-methylbutylmalonyl-CoA
Abs260nm (mAU)
time (min)

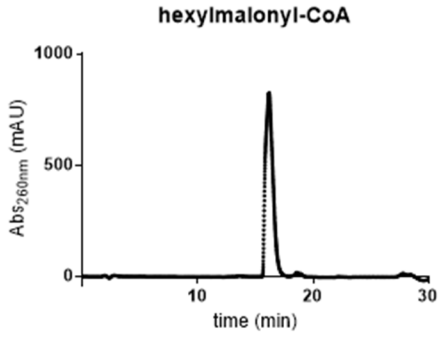

3-methylbutylmalonyl-CoA
relative abundance [%]
m/z

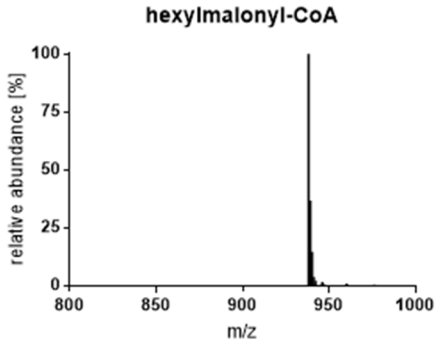

benzylmalonyl-CoA
Abs260nm (mAU)
time (min)


benzylmalonyl-CoA
relative abundance [%]
m/z


butylmalonyl-CoA
Abs260nm (mAU)
time (min)


butylmalonyl-CoA
relative abundance [%]
m/z


hexylmalonyl-CoA
Abs260nm (mAU)
time (min)


hexylmalonyl-CoA
relative abundance [%]
m/z

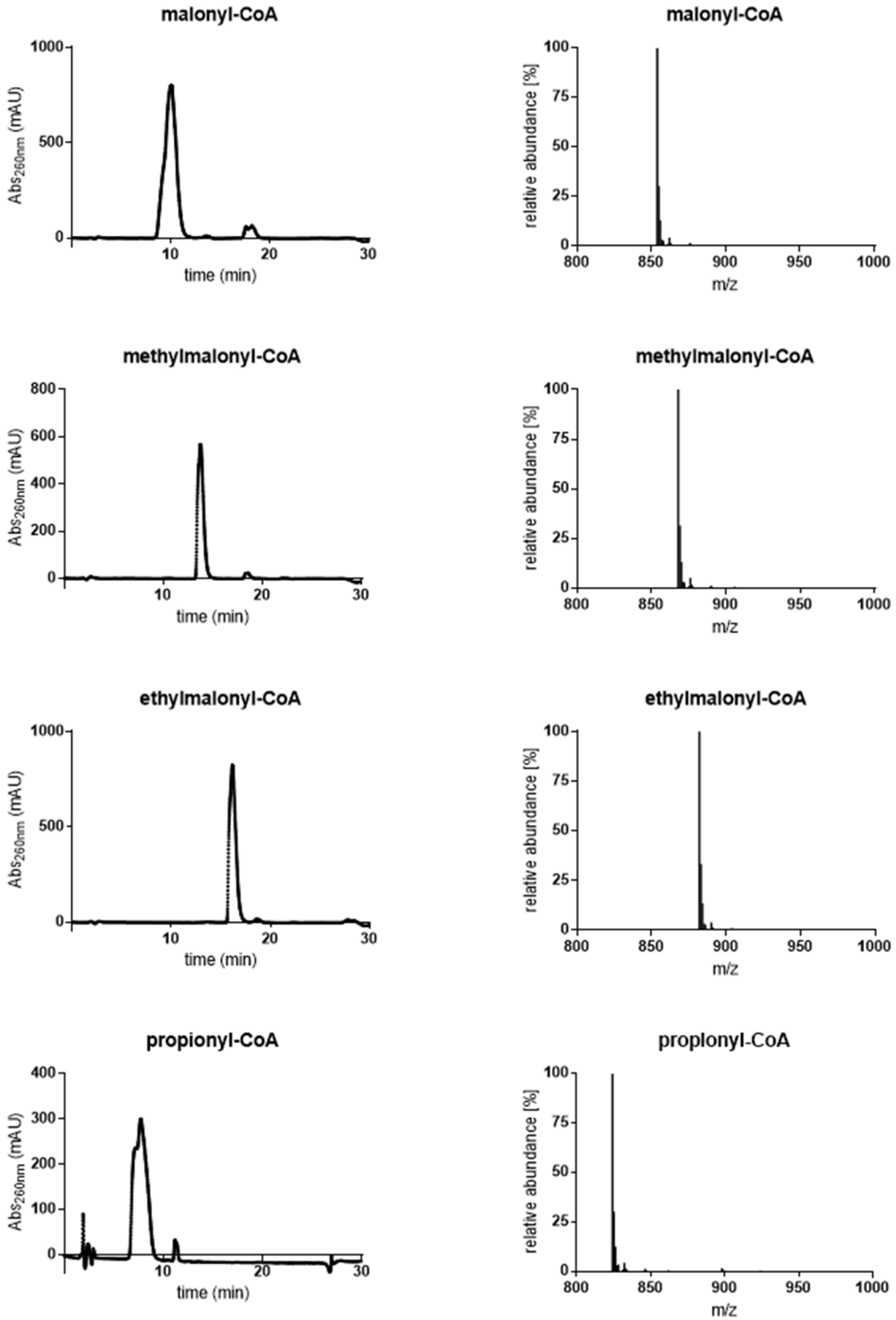


Figure S4 (related to Figure 3): Purity of all synthesized extender units assessed by HPLC-MS (left panel; HPLC trace at 260 nm, right panel; *m/z* of the elution peak)

Experimental section

Key Resources Table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial Strains		
E.coli BAP1	[3]	N/A
E.coli BL21-AI	Invitrogen	Cat#C607003
E.coli DH5α	NEB	Cat#C2987I
Chemicals		
NADPH	Biomol	Cat#16156.1
CoA	Roche	Cat#10155969
Crotonic anhydride	Sigma	Cat#130974
Propionic anhydride	Sigma	Cat#240311
Hexanoic acid	Sigma	Cat#153745
Octanoic acid	Sigma	Cat#C2875
5-methylhexanoic acid	Sigma	Cat#CDS003381
trans-Cinnamic acid	Sigma	Cat#C80857
3-phenylpropionic acid	Sigma	Cat#135232
CDI	Sigma	Cat#115533
Triethylamine	Sigma	Cat#T0886
MeOH HPLC grade	Roth	Cat#KK39.2
MeCN HPLC grade	Roth	Cat#HN44.2
Ammonium formate	Fluka	Cat#17843
Kaliumhydrogencarbonate	Roth	Cat#P748.2
Tetrahydrofuran	Roth	Cat#4745.2
Ethylchloroformate	Sigma	Cat#185892
DMF	Sigma	Cat#227056
DCM	Sigma	Cat#270997
Tris	Roth	Cat#4855.2
Protino Ni-NTA agarose	Macherey-Nagel	Cat#1709
Recombinant Proteins		
pET16b-$CcrC_{PAG}$	[1]	see Table S4
pET16b-AntE	this work	see Table S4
pET16b-Ndas_0488	this work	see Table S4
pET16b-KSE_65530	this work	see Table S4
pET16b-DivR	this work	see Table S4
pET16b-Acx4	[4]	see Table S4
pET28b-LD(4) (also pBL12)	[5]	N/A
pET28b-Strep-LD(4)	this work	N/A
pET28b-(5)Mod1(2) (also pBL13)	[5]	N/A
pET28b-(3)Mod2TE (also pBL16)	[5]	N/A
Software		
eMZed	[6]	http://emzed.ethz.ch/
GraphPad Prism 7	GraphPad Software	https://www.graphpad.com/
OpenLAB CDS ChemStation	Agilent Software	https://www.agilent.com/
MassHunter Software	Agilent Software	https://www.agilent.com/
CaryWin UV Software	Agilent Software	https://www.agilent.com/

Bacterial Strains and Cultivation. All used bacterial strains are listed in the Key Resources Table. *E.coli* DH5*α* was used for cloning and site-directed mutagenesis. *E.coli* BL21-AI was used for the expression of ECRs and Acx4 and *E.coli* BAP1 was used for expression of DEBS. *E.coli* was routinely grown in LB at 37°C, protein expressions were done in TB at the temperatures described in the Method Details.

Plasmid Construction. The expression plasmid for CinF was a generous gift from Rolf Müller[2]. The expression plasmid for $CcrC_{PAG}$ was previously cloned into pTE-16b using NdeI and BamHI, mutations were introduced using the QuickChange Site-Directed Mutagenesis Kit (Stratagene, La Jolla, USA)[1]. The antE, divR, Ndas_0488 and KSE_65530 were codon optimized and synthesized by the DOE Joint Genome Intitute adding an N-terminal 10x His tag and cloned into the backbone pET16b using NcoI and XhoI (see Table S4 for complete DNA and protein sequences).

ECR and Acx4 preparation. *E.coli* BL21 (DE3) (for Acx4 *E.coli* BL21 AI) were transformed with the respective expression plasmid and plated on LB-Agar containing 100 µg/ml Ampicilin (50 µg/ml Kanamycin for CinF). Pooled colonies of overnight inoculated petri dishes were used to inoculate 1 L of TB containing the respective antibiotic. The expression culture was incubated at 37°C until an OD_{600} of 0.8 was reached. Cultures for ECR expressions were cooled down to 20°C and induced with 0.25 mM IPTG and incubated overnight. Acx4 expression cultures were cooled to 30°C and induced with 0.25 mM IPTG and 2‰ L-arabinose and incubated for four hours. Cells were harvested, resuspended in 50 ml Buffer A (50 mM TrisHCl pH 7.9, 500 mM NaCl) and lysed by sonication. Lysates were centrifuged at 42,000 g at 4°C for 45 min. The supernatant was then loaded onto a pre-equilibrated 1 mL HisTrap column (GE Healthcare) and washed with 15% Buffer B (50 mM TrisHCl pH 7.9, 500 mM NaCl, 500 mM imidazole) and eluted with 100% Buffer B. The buffer of the purified proteins was exchanged into the protein storage buffer (30% glycerol, 150 mM NaCl, 25 mM TrisHCl pH 7.9). Protein purity was checked by SDS-gel electrophoresis; concentration of proteins was determined spectroscopically at 280 nm using calculated extinction coefficients. For

Acx4 the FAD concentration was determined at 450 nm and FAD was added to reach a equimolar protein to FAD ratio. Proteins were stored at -20°C.

DEBS protein preparation. The expression plasmids for the *in vitro* triketide model system of DEBS were constructed according to a previously described protocol yielding the plasmids pBL12 (here termed pET28b-LD(4); the N-terminal HisTag of pBL12 was exchanged for a StrepTag), pBL13 (pET28b-(5)Mod1(2)) and pBL16 (pET28b-(3)Mod2TE)[5]. *E. coli* BAP1 was transformed with expression plasmid, 1 L of TB expression culture was grown to an OD_{600} of 0.7, cooled down to 18°C, induced with 0.25 mM IPTG and incubated over night at 18°C. Cells producing module 1 and module 2 were harvested, resuspended in Buffer C (50 mM NaH_2PO_4 pH 7.6, 450 mM NaCl, 20% glycerol), lysed by sonification and centrifuged at 42,000 g at 4°C for 45 min. The supernatant was incubated with pre-equilibrated Protino Ni-NTA beads (1 mL resin) for 1 h. The beads were washed with 5% Buffer D (75 mM NaH_2PO_4 pH 7.6, 500 mM imidazole, 20 mM NaCl, 10% glycerol) and then eluted with 100%. The elution was loaded onto a with Buffer E (50 mM NaH_2PO_4 pH 7.6, 10% glycerol) pre-equilibrated HiTrap Q anion exchange column and eluted in a 50 mL NaCl gradient with a final concentration of 500 mM NaCl. Cells expressing the plasmid yielding the loading domain were resuspended in Buffer C, processed as described above and purified over a pre-equilibrated 1 mL StrepTrap column (GE Healthcare), using 2 mM d-Desthiobiotin (Sigma-Aldrich) for elution. Fractions were tested on an SDS-PAGE gel, pooled and concentrated with an Amicon Ultra-4 100 kDa (50 kDa for LD(4)) centrifugal filter. A final concentration of 20% glycerol was added to the concentrated proteins, their concentration was determined spectroscopically at 280 nm and they were stored at -80°C.

Chemical CDI Coupling. CDI (1,1′-carbonyldiimidazole) coupling of carboxylic acids to CoA was done according to a previously described protocol[7]. 42 mg CDI (0.27 mmol, 4 eq.) was dissolved in 2 mL THF and the acid was added (0.31 mmol, 4.8 eq.; hexanoic acid: 38.9 µl, octanoic acid: 49.2 µl, dihydrocinnamic acid: 46.5 µl and 5-methylhexanoic acid: 43,5 µl). The reactions were stirred for 1 h at 22°C. 50 mg CoA was dissolved in 500 µL 1 mL $KHCO_3$ and added to the reaction. The reaction was stirred for another

45 min and tested for remaining free thiols using DTNB, then flash frozen in liquid N2 and lyophilized overnight. Samples were then dissolved in 1 mL H2O and used for HPLC-purification or used for desaturation by Acx4 or for direct one-pot biosynthesis of the extender units.

Synthesis of cinnamoyl-CoA using ethylchloroformate coupling. Synthesis of cinnamoyl-CoA was done by chemically coupling cinnamic acid to CoA according to a previously described protocol[7]. 37 mg trans-cinnamic acid (0.25 mmol, 2 eq.) was dissolved in 2 mL DCM, 37 µl trimethylamine (0.26 mmol, 2.1 eq) was added and the reaction was stirred at 20°C for 30 min. The mixture was cooled on ice and 24 µl ethylchloroformate (0.25 mmol, 2 eq.) was added. The mixture was stirred on ice for 2 h. The DCM was evaporated at room temperature and resuspended in 2 mL DMF. 100 mg CoA (0.125 mmol, 1 eq.) was dissolved in 300 µL 1 M $KHCO_3$ and added to the mixture at 20°C. Completion of the reaction was tested using DTNB, the reaction mixture was quenched by adding 500 µL of 50% formic acid, frozen in liquid N_2 and lyophilized overnight. The sample was dissolved in 2 mL H_2O and used for HPLC purification.

Synthesis of malonyl- and methylmalonyl-CoA. Synthesis of malonyl- and methylmalonyl-CoA was done according to a previously described protocol[7]. 20 mg CoA (1 eq.), 13.2 mg malonic acid (5 eq.) or 15.2 mg methylmalonic acid (5 eq.) and 70.4 mg ATP (4 eq.) were dissolved in 5 mL of 200 mM $KHCO_3$ containing 15 mM $MgCl_2$ and 3.2 µM MatB. The reaction was incubated at 30°C and completion confirmed with DTNB. The mixture was quenched with 5% formic acid and directly used for HPLC purification.

Synthesis of ethylmalonyl-CoA. Ethylmalonyl-CoA was synthesized by chemical coupling of crotonic anhydride with CoA followed by the addition of a $CcrC_{PAG}$, NADPH and $KHCO_3$. 10 mg CoA was dissolved in 1 mL 1 M $KHCO_3$.The mixture was cooled on ice and 3.2 µl of crotonic anhydride was added and stirred on ice for 30 min. Completion of the reaction was tested using DTNB. The reaction mixture was then added to 8 mL of 250 mM TrisHCl pH 7.5 containing 30 mM NADPH, 250 mM $KHCO_3$ and 100 nM

$CcrC_{PAG}$. The reaction was incubated for 30 min at 30°C, quenched with final concentration 5% formic acid and directly used for HPLC purification.

One pot biosynthesis of extender units. The enzymatic desaturation with Acx4 as well as the reductive carboxylation was then done using this lyophilized reaction mixture without any purification in between. The lyophilized CDI reaction mixture (see above) was resuspended in 1 mL of H_2O and used directly for the biosynthesis of the extender units. A 10 mL assay for extender unit biosynthesis contained 2 mL of 0.5 M TrisHCl pH 7.5 (final concentration 100 mM), 1 mL of 1 M $KHCO_3$ (100 mM), 500 µL of 200 mM NADPH (10 mM), 300 µl of the dissolved CDI reaction mixture (theoretical maximum of 50 mM, resulting in an approximate final concentration of 1 mM in the assay), 500 µl of 48 µM Acx4 (2.4 µM) and 200 µl of 39 µM $CcrC_{PAG}$ (0.8 µM). The mixture was shaken at 200 rpm in a baffled Erlenmeyer flask at 30°C for 120 min. To follow progress of the reaction, 30 µL of the reaction mixture were quenched at various time points with the addition of 3 µL of 50% formic acid and analyzed by analytical HPLC-MS (Figure 3). The reaction was quenched after 120 min with the addition of 1 mL of 50% formic acid, centrifuged for 10 min at 17000 g to remove precipitated protein and then purified via HPLC. The purified product was lyophilized, resuspended in water and the final yield of the overall reaction and purification was determined by determining the extender unit concentration via UV/Vis absorption at 260 nm (calculated as (mol product)/(mol CoA) Table S3).

HPLC purification of acyl-CoA esters. All synthesized CoA-thioesters were purified using a 1260 Infinity LC system (Agilent) using a Gemini 10 µm NX-C18 110 Å, 100 x 21.2 mm, AXOA packed column (Phenomenex). The general protocol used a flow rate of 25 ml min^{-1} starting with 5 min of 5% MeOH in 50 mM NH_4HCO_2 pH 8.2, followed by a gradient from 5% to 40% MeOH in 15 min, a 2 min washing step at 95% MeOH and a re-equilibration step of 3 min at 5% MeOH. Purified CoA-esters were lyophilized and the purity was checked using analytical HPLC-MS according to the protocol described for the carboxylation vs. reduction assays.

Determination of carboxylation vs. reduction ratio of ECR variants. Assays for the determination of carboxylaton vs. reduction contained 100 mM TrisHCl pH 8, 100 mM $KHCO_3$, 10 µg mL^{-1} carbonic anhydrase, 150 µM NADPH, 75 µM substrate and 1.5 µg of the respective ECR. The reaction was followed on a Cary-60 UV/Vis spectrometer and quenched upon completion with final concentration 5% formic acid. Assays were then analyzed on a 1260 Infinity LC system (Agilent) with attached with a 6130 Quadrupole LC/MS detector (Agilent) using a Luna 3 µm C18(2) 100 Å, 150 x 2 mm (Phenomenex) column with a flow rate of 0.3 ml min^{-1}. For malonyl-, methylmalonyl- and ethylmalonyl-CoA a gradient starting with 3 min of 5% MeOH in 50 mM NH_4HCO_2 pH 8.2, followed by a gradient from 5% to 25% MeOH over 20 min, a 2 min washing step at 95% MeOH and a 3 min re-equilibration step at 5% MeOH. All other samples were run using a gradient starting with 3 min of 5% MeCN in 50 mM NH_4HCO_2 pH 8.2, followed by a gradient from 5% to 35% MeCN over 20 min, a 2 min washing step at 95% MeCN and a 3 min re-equilibration step at 5% MeCN.

Spectrophotometric enzyme assays. Enzyme assays were carried out on a Cary-60 UV/Vis spectrometer (Agilent) using quartz cuvettes (1 mm and 10 mm diameter, Hellma). Acx4 characterization was done using 10 mm cuvettes in 100 mM TrisHCl pH 8 buffer following double bond formation at 290 nm ($\Delta\varepsilon290$ = 2.26 $mM^{-1}cm^{-1}$) for all substrates except for phenylpropionyl-CoA (dihydrocinnamoyl-CoA), which was measured at 308 nm ($\Delta\varepsilon308.$=16.4 $mM^{-1}cm^{-1}$, Figure S2B). ECR characterization for all substrates was in 10 mm cuvettes following NADPH consumption at 360 nm in 100 mM TrisHCl pH8 buffer containing 100 mM $KHCO_3$ and 500 µM NADPH (Figure S4).

Extender unit incorporation assays. Competition assays to test incorporation of extender units into triketides were run at 30°C in 200 mM NaH_2PO_4 with 2 mM NADPH, 4 µM epimerase, 0.2 mM propionyl-CoA, 1 mM of each extender unit and 1 µM of LD(4), (5)Mod1(2) and one of the (3)Mod2TE variants. Assays were quenched with 5% formic acid after 60 min and analyzed via HPLC-ESI-TOF on a 6550 iFunnel Q-TOF LC-MS (Agilent) using a 1.8 µm Zorbax SB-C18 column, 50 x 2.1 mm (Agilent) using H_2O (A) and MeCN (B) both containing 0.1% formic acid. The gradient condition were as follows:

0 min 5% B; 1 min 5% B; 6 min 95% B,6.5 min 95% B, 7 min 5% B with a flow rate of 250 µl min^{-1}. Capillary voltage was set at 3.5 kV and nitrogen gas was used as nebulizing (20 psig), drying (13 L^{-1}min,225°C) and sheath gas (12 L min^{-1}, 40°C). MS data were acquired with a scan range of 100-1000 m/z. Data were analyzed using MassHunter Qualitative Analysis software (Agilent) and eMZed[6].

Quantification and statistical analysis. For all Michaelis-Menten plots to determine the kinetic parameters of ECRs and Acx4 at least 18 points were measured. The error is given as standard deviation (Figure S1 and Table S2). Carboxylation vs. reduction ratios were determined in triplicates (Figure 1). One pot extender unit biosynthesis was carried out multiple times and a representative progress curve of one assay is shown in Figures 2 and 3, yields of Table S3 were determined for the shown traces. The *in vitro* competition assays were carried out twice with independently expressed and purified protein batches.

References

1. Peter, D. M.; Schada von Borzyskowski, L.; Kiefer, P.; Christen, P.; Vorholt, J. A.; Erb, T. J., Screening and engineering the synthetic potential of carboxylating reductases from central metabolism and polyketide biosynthesis. *Angewandte Chemie International Edition* **2015,** *54* (45), 13457-13461.
2. Quade, N.; Huo, L.; Rachid, S.; Heinz, D. W.; Müller, R., Unusual carbon fixation gives rise to diverse polyketide extender units. *Nature chemical biology* **2012,** *8* (1), 117-124.
3. Pfeifer, B. A.; Admiraal, S. J.; Gramajo, H.; Cane, D. E.; Khosla, C., Biosynthesis of complex polyketides in a metabolically engineered strain of E. coli. *Science* **2001,** *291* (5509), 1790-2.
4. Schwander, T.; von Borzyskowski, L. S.; Burgener, S.; Cortina, N. S.; Erb, T. J., A synthetic pathway for the fixation of carbon dioxide in vitro. *Science* **2016,** *354* (6314), 900-904.
5. Lowry, B.; Robbins, T.; Weng, C. H.; O'Brien, R. V.; Cane, D. E.; Khosla, C., In vitro reconstitution and analysis of the 6-deoxyerythronolide B synthase. *J Am Chem Soc* **2013,** *135* (45), 16809-12.
6. Kiefer, P.; Schmitt, U.; Vorholt, J. A., eMZed: an open source framework in Python for rapid and interactive development of LC/MS data analysis workflows. *Bioinformatics* **2013,** *29* (7), 963-964.
7. Peter, D. M.; Vögeli, B.; Cortina, N. S.; Erb, T. J., A chemo-enzymatic road map to the synthesis of CoA esters. *Molecules* **2016,** *21* (4), 517.

Chapter II

Understanding Substrate Selectivity of Phoslactomycin Polyketide Synthase by Using Reconstituted *in vitro* Systems

Published in: ChemBioChem 2020, 21, 1-7

Understanding substrate selectivity of phoslactomycin polyketide synthase using reconstituted in vitro systems

Kyra Geyer[a], Srividhya Sundaram[a], Peter Sušnik[b], Ulrich Koert[b], Tobias J. Erb*[a, c]

[a]Department of Biochemistry and Synthetic Metabolism; Max Planck Institute for Terrestrial Microbiology, Karl-von-Frisch-Str. 10, D-35043 Marburg, Germany, toerb@mpi-marburg.mpg.de [b]Department of Chemistry, Philipps-University Marburg, Hans-Meerwein-Str. 4, D-35032 Marburg, Germany [c]LOEWE Center for Synthetic Microbiology (Synmikro), Karl-von-Frisch-Str. 16, D-35043 Marburg, Germany. *Corresponding author: toerb@mpi-marburg.mpg.de

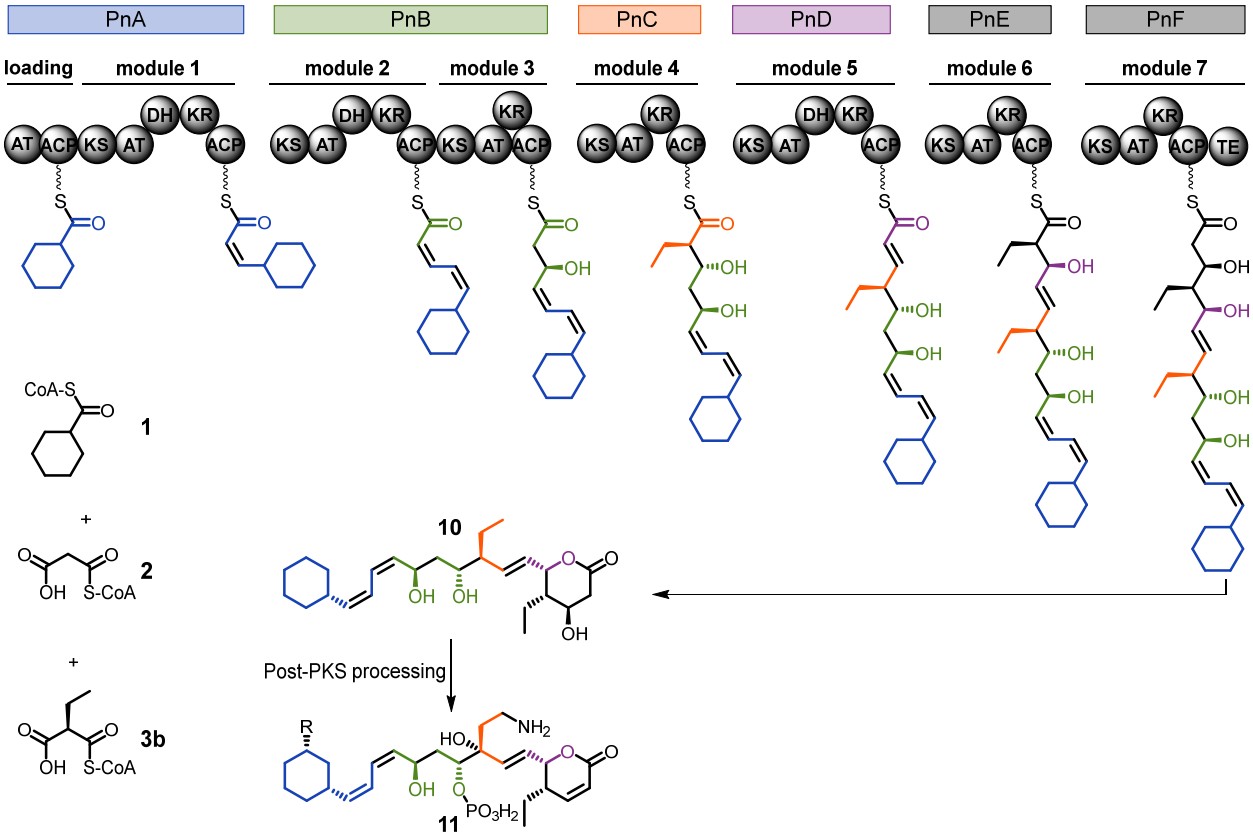
ChemBioChem 2020, 21, 1-7

Author contributions: K.G. and T.J.E conceived the project. K.G., designed and performed experiments and analyzed the data. S.S performed expression experiments. P.S. synthesized the SNAC ester derivatives. K.G. and T.J.E. wrote the manuscript with contributions from all authors.

Abstract: Polyketide synthases (PKS) use simple extender units to synthesize complex natural products. A fundamental question is how different extender units are site-specifically incorporated into the growing polyketide. Here we established phoslactomycin (Pn) PKS that incorporates malonyl- and ethylmalonyl CoA as an *in vitro* model to study substrate specificity. We combined up to six Pn PKS modules with different termination sites for the controlled release of tetra-, penta- and hexaketides and challenged these systems with up to seven different extender units in competitive assays to test for specificity of Pn modules. While malonyl-CoA modules of Pn PKS exclusively accept their natural substrate, the ethylmalonyl-CoA module PnC tolerates different *α*-substituted derivatives, but discriminates against malonyl CoA. We show that the ratio of extender transacylation to hydrolysis controls incorporation in PnC, explaining site-specific selectivity and promiscuity in the natural context of Pn PKS.

Introduction

Polyketides are a class of natural products with high structural diversity but common biosynthetic logic. They are synthesized by polyketide synthases (PKSs) from simple acyl and (alkyl)malonyl CoA-based building blocks. The structure of the product is determined by the type and location of catalytic domains in the assembly line, while structural diversity of the backbone arises from the choice of extender unit, extent of reduction and stereo centers installed during biosynthesis [1, 2]. The three major extender units incorporated in polyketides are malonyl-, methylmalonyl-CoA and to a lesser extent ethylmalonyl-CoA. While it is generally believed that the acyltransferase (AT) domains control extender unit incorporation, recent reports have shown that these domains are more promiscuous than initially assumed. For example AT5 of monensin accepts methyl- and ethylmalonyl-CoA in the natural system, but experiments showed that also propargyl- and butylmalonyl-CoA are accepted [3, 4]. Research on the 6-deoxyerythronolide B PKS (DEBS), a well-established model system, revealed an intrinsic promiscuity of the ATs of module 2, module 3 [5] and module 6 [6]. Promiscuous AT domains were also observed in the structurally highly similar pikromycin PKS (module 5 and module 6 [7, 8]). However, to fundamentally understand how extender unit selectivity in PKS is achieved other, more diverse model systems are required. Especially, since the well-studied DEBS PKS does not need to distinguish between different extender units, because only methylmalonyl-CoA is incorporated into the assembly line. Here, we focused on phoslactomycin (Pn) PKS, a modular type I PKS from *Streptomyces platensis* [9, 10]. Pn PKS is composed of a loading module selecting cyclohexanecarboxyl-CoA (**1**) and seven extending modules of which five incorporate malonyl-CoA (**2**) and two incorporate ethylmalonyl-CoA (**3b**) (Scheme 1). This makes Pn PKS, in comparison to methylmalonyl-CoA specific DEBS PKS, an excellent model to study how extender unit selectivity is controlled at different sites within one PKS.



Scheme 1. Phoslactomycin polyketide synthase Pn PKS. ACP=acyl carrier protein; AT=acyltransferase; DH=dehydratase; KR=ketoreductase; TE=thioesterase; **1**=cyclohexanecarboxyl-CoA; **2**=malonyl-CoA; **3**=(2*S*)-ethylmalonyl-CoA; **10**=phoslactomycin polyketide backbone, **11**=bioactive phoslactomycin derivatives; **R**=isobutyloxy; isovaleryloxy; 4-methylcaleryloxy; cyclohexylcarbonyloxy; 4-methylheptanoyloxy [9, 10].

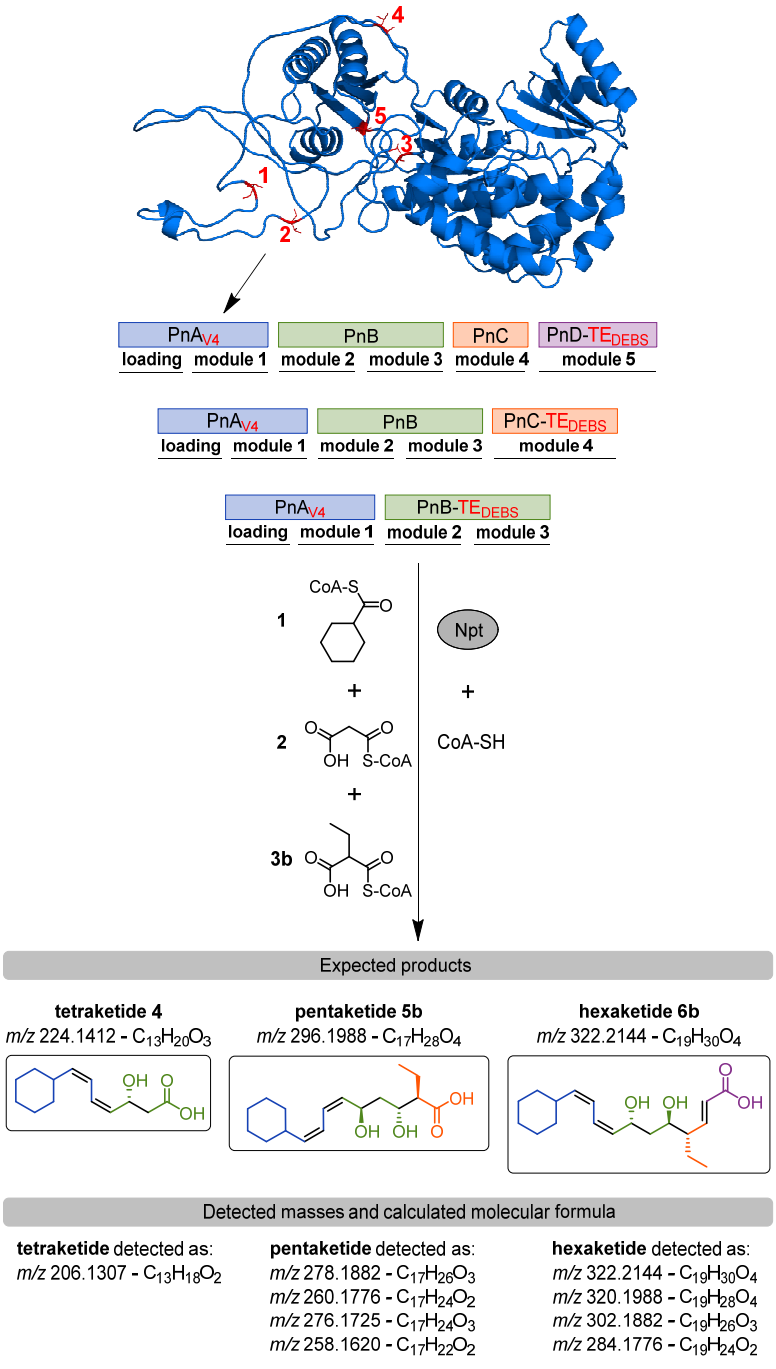
Results and Discussion

We aimed at establishing different Pn-based minimal model systems for the production of tetra-, penta- and hexaketides to study promiscuity and selectivity, in particular of the **3b** incorporating PnC_AT. To that end, we expressed PnA-PnD as individual proteins in *E. coli* BAP1 [7, 8, 11, 12] (Figure S1). For PnA production, we tested five different transcription start sites (PnA_{V1}-PnA_{V5}, Scheme 2), due to insolubility upon expression of the full-length annotated PnA protein. Variant PnA_{V4} resulted in highest protein expression and solubility. To enable polyketide release in the tetra-, penta- and hexaketide systems, we used the chain releasing thioesterase (TE) from DEBS to create chimeric PnB-TE_{DEBS}, PnC-TE_{DEBS} and PnD-TE_{DEBS} constructs [8, 13, 14] (DNA sequence information in SI).

E. coli BAP1 encodes for a functional phosphopantetheinyl (Ppant) transferase. Yet, phosphopantetheinylation highly varied across individual Pn acetyl-carrier proteins (ACPs) (Table S1). While PnC_ACP was fully converted to *holo*-ACP (96%), only 26% of PnB_ACP1 was present in the *holo*-form. Analysis revealed one gene from the 4'-phosphopantetheinyl transferase superfamily (Npt, OSY40025) with low homology to Sfp (16.1% sequence identity) in the genome of *S. platensis*. We expressed the protein in *E. coli* BL21 (DE3) and demonstrated that Npt serves as a bona fide Ppant transferase that is able to activate *apo*-Pn_ACPs (Figure S2). We further added purified Npt to all *in vitro* assays to ensure the full conversion of all Pn proteins to their *holo*-form.

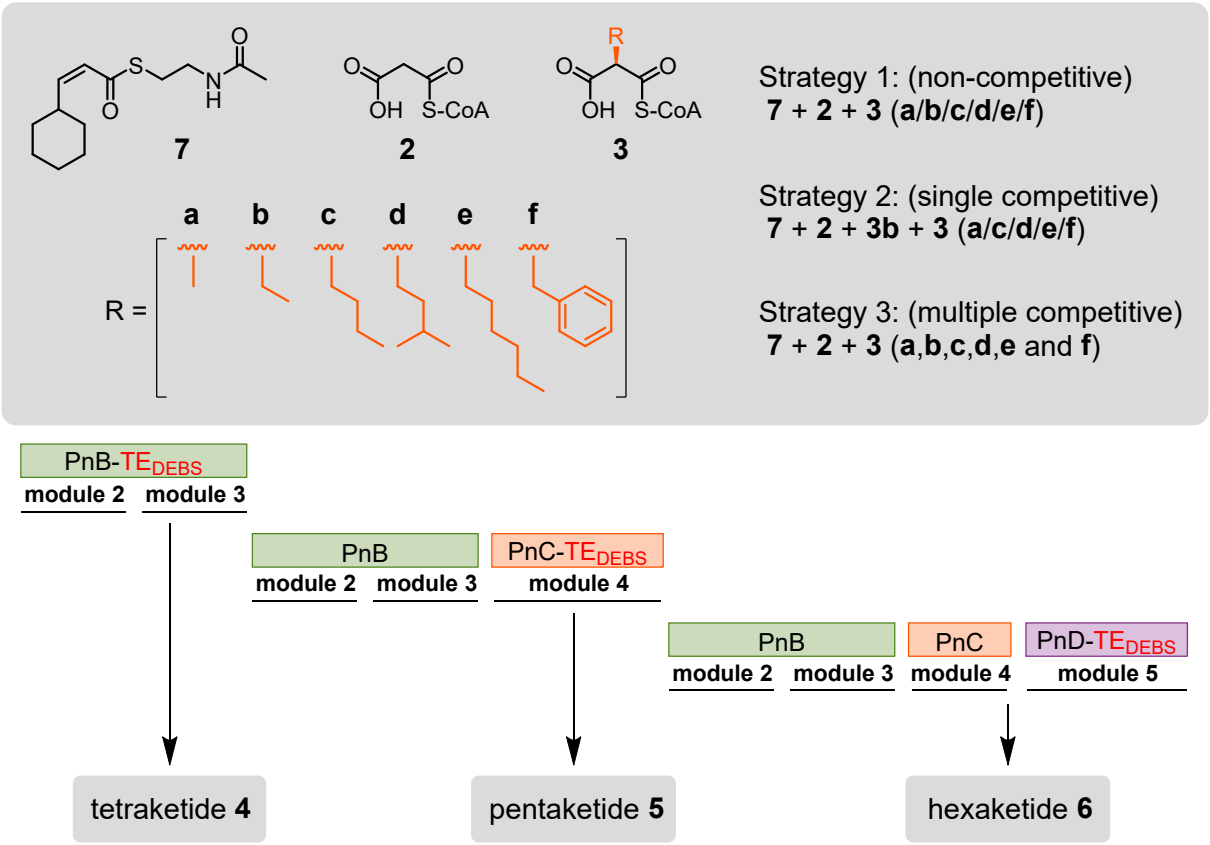
One-pot reaction mixtures of PnA_{V4} and PnB-TE_{DEBS} (tetraketide system), PnA_{V4}, PnB and PnC-TE_{DEBS} (pentaketide system), PnA_{V4}, PnB, PnC and PnD-TE_{DEBS} (hexaketide system), incubated with their natural substrates resulted in production of respective polyketide derivatives, as analyzed by high resolution LC-MS. Enzymatic production of these derivatives was additionally verified using 2-C^{13} labelled malonyl-CoA (Figure S3, Figure S4 and Figure S5). A product corresponding to a dehydrated tetraketide (*m/z* 206.1307, formula $C_{13}H_{18}O_2$) is the exclusive product found in tetraketide assays. In pentaketide assays, a more complex product spectrum was observed, with *m/z* 258.1620 corresponding to a pentaketide of formula $C_{17}H_{22}O_2$ as the main product. Various masses were found in hexaketide assays that correspond to different products with *m/z* 284.1776 corresponding to hexaketide of formula $C_{19}H_{24}O_2$ as the main product

(Scheme 2). Based on the combined information from the high resolution mass spectrometer and isotopic labeling, we propose structures for the individual masses detected, and provide an explanation for their generation in Figure S6. Altogether, our analysis demonstrated that the Pn PKS enzymes are functional *in vitro* and can be combined stepwise to produce different truncated Pn polyketide variants (Scheme 2 and Figure S6)



Scheme 2. *In vitro* reconstitution of Pn PKS. Structural model of PnA loading AT with the alternative transcription start sites V1 to V5 highlighted in red. *S. platensis* endogenous Ppant transferase Npt and CoA-SH were used to ensure full conversion into the corresponding *holo*-PKS *in vitro*. **1**, **2** and **3b** are the natural substrates of Pn PKS. Termination of the polyketide production with PnB-TE_{DEBS}, PnC-TE_{DEBS} or PnD-TE_{DEBS} is expected to yield the depicted tetra- (**4**), penta- (**5b**) or hexaketide (**6b**). Mass spectrometric analysis detected derivatives thereof indicating dehydrations and oxidations. A detailed discussion of the individual masses, the corresponding putative products and their generation is provided in Figure S6.

Successful reconstruction of the first six modules of Pn PKS allowed us to assess the substrate tolerance of the system with seven different extender units (**2**, **3a**-**f**). In these assays, we bypassed PnA_{V4} by using the *N*-acetylcysteamine (SNAC) analogue of the diketide product (7, see SI for synthesis). We tested substrate tolerance across the tetra-, penta- and hexaketide systems employing three different strategies. In strategy 1, we used **7**, **2** (malonyl-CoA) and one of the six *α*-substituted extender units (**3a**-**3f**) to test if substrates other than ethylmalonyl-CoA (**3b**) are incorporated. In strategy 2, we used **7**, **2**, **3b** and one additional extender unit (**3a**, **3c**-**3f**), to set the incorporation of **3b** in direct competition to an alternative non-native extender unit. In strategy 3, we used **7**, **2**, and all six *α*-substituted extender units in parallel (**3a**-**f**) to allow identification of the preferred substrate(s) (Scheme 3).



Scheme 3. Pn PKS *in vitro* characterization. **7** diketide product of PnA_{V4} as SNAC thioester analogue, **2** malonyl-CoA, **3** (*2S*)-acyl-malonyl-CoA derivatives with residues marked in orange: **a** methyl-, **b** ethyl-, **c** butyl-, **d** 3-methylbutyl-, **e** hexyl-, **f** benzylmalonyl-CoA. Polyketide production is terminated with PnB-TE_{DEBS}, PnC-TE_{DEBS} or PnD-TE_{DEBS} to produce **4**, **5** and **6**, respectively.

Using strategy 1 and strategy 2 with the tetraketide system, we observed exclusive incorporation of 2, as indicated by product analysis. No alternative extender unit was accepted, indicating exquisite selectivity by PnB. In the pentaketide system, incorporation of all six *α*-substituted extender units **3a-3f** was observed, while **2** was not incorporated at the position of PnC (Figure S4). This demonstrates substrate tolerance of PnC towards *α*-substituted extender units, but at the same time strong discrimination against **2**. More specifically, similar amounts of **5b**, **5c**, **5d** and **5e** were detected in the pentaketide system, when testing substrate tolerance with strategy 1 (Figure 1A). Products of **3a** incorporation showed less intensity and products of **3f** incorporation were detectable in traces. In strategy 2, when we put incorporation of **3b** (natural substrate) in direct competition to another extender unit, we again observed tolerance of PnC towards

medium chain extender units (**3b-d**), and discrimination against short (**3a**) and long chain (**3e** and **3f**) extender units. With strategy 3, the trend increased further towards preferred incorporation of **3b** (natural substrate) and **3c** by the pentaketide system.

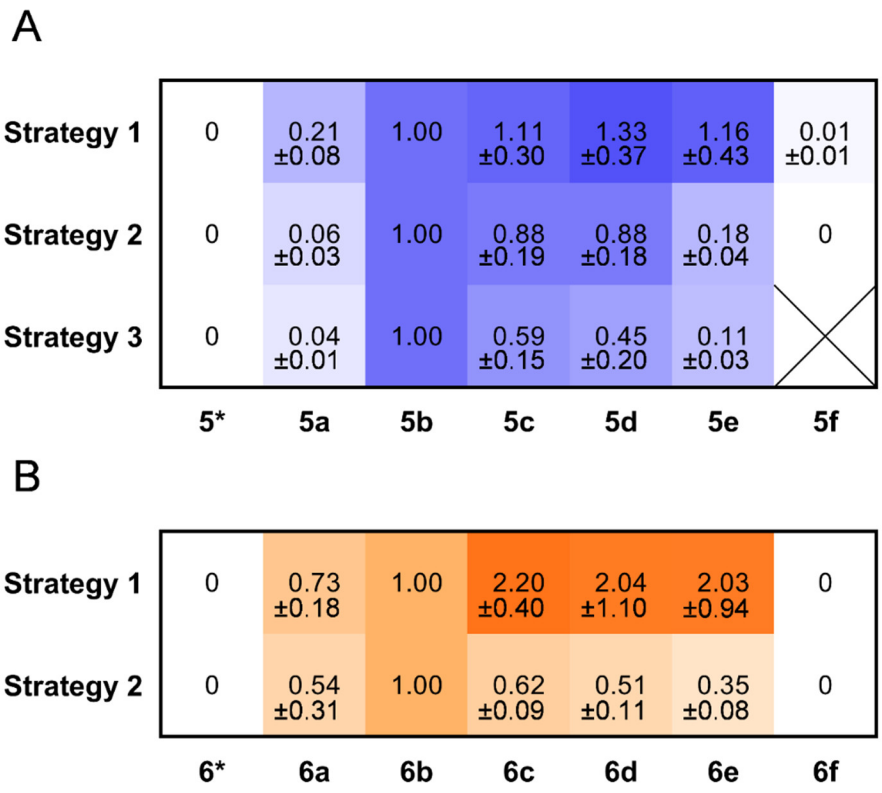


Figure 1. Relative amounts of pentaketides and hexaketides in competition assays. Ion count of each product is set relative to **5b** or **6b**. **5*** or **6*** indicate incorporation of **2** by PnC. Given is the mean of minimum three biological replicates, with each two technical replicates (pentaketide assays) and the mean of three biological replicates, with each two technical replicates (hexaketide assays).

Having demonstrated that PnC contains a highly promiscuous AT and is able to form different pentaketides by incorporation of **3a-3f**, we asked, whether PnD is able to accept and process the unnatural pentaketides. We used strategy 1 and 2 to study substrate tolerance of the reconstituted hexaketide system. When only one *α*-substituted substrate is available, **6a-6e** but not **6f** were formed (Figure 1B, Figure S5). Also, we did not detect multiple incorporations of non-native extender units, indicating that PnC is the only entry point of non-native extender units. With strategy 2, natural product **6b** became the main product. This demonstrated that Pn PKS is generally flexible for incorporation of unnatural

extender units at the level of PnC (Figure 1A, strategy 1) but that in presence of the natural substrate a preference for the latter exists (Figure 1B, strategy 2).

When comparing the product profile of strategy 2 to strategy 1 in hexaketide assays, we observed that relative product formation decreased with increasing size of the side chain. For **6a** we observed a decrease of only 26%, whereas for **6c**, **6d** and **6e** the relative decrease increased between 72% and 83%. Furthermore, compared to **5a**-**5b** and **6a**-**6b** that were detected as reduced product, alongside oxidized products (Figure S6, Table S2), **5c**-**5f** and **6c**-**6e** were exclusively found as the non-reduced products, indicating the polyketide is not fully processed by subsequent domains after incorporation of the non-native, longer extender unit, which is in line with earlier reports [8]. Taken together this indicates a selection by PnD for pentaketides with shorter site chains, either due to restrictions in the KS active site and/or by preferred reduction (and further processing) of those pentaketides.

Next, we wanted to study the specificity of acyl transfer at single domain level. To that end, we aimed at measuring kinetic parameters for transacylation and hydrolysis in ketosynthase-AT (KS-AT) didomains. In the commonly used assay, these kinetic data are collected by coupling CoA-SH release by the AT domain to production of succinyl-CoA by a commercially available *α*-ketoglutarate dehydrogenase complex (*α*-KGDH) [12, 15, 16]. However, we observed varying apparent (app.) AT activity that was dependent on the *α*-KGDH batch used. We therefore developed a steady state kinetic assay based on purified *E. coli* succinyl-CoA ligase complex SucC/SucD, to minimize AT-inhibition (Scheme S1). In this assay, when the acyl-residue is transferred to the AT, free CoA-SH is released. The free CoA-SH is coupled to succinate via SucC/D, whereby ADP is released. ADP production is subsequently coupled to NADH consumption via pyruvate kinase/lactate dehydrogenase. Thus, the release of CoA-SH is coupled to NADH consumption, which can be spectrophotometrically measured and quantified.

With the SucC/SucD-based assay we determined kinetic parameters of PnB_KS-AT1, PnB_KS-AT2, PnC_KS-AT and PnD_KS-AT for hydrolysis and transacylation (Table S3, Figure S7). As negative control AT catalytic knockouts were used. All ATs described to incorporate **2** (PnB module 2, module 3 and PnD), showed hydrolysis and transacylation activity only for their natural substrate, which is in line with

our *in vitro* experiments in the reconstituted system. These measurements were independently confirmed by mass spectrometric analysis of acylated ACPs, in which we exclusively detected malonyl-ACP formation (Table S1, Table S4). All three ATs showed comparable hydrolysis rates at an app. k_{cat} of approx. 3 min^{-1} and an app. K_M between 4 µM to 11 µM. Rates for transacylation were between 10- to 35- fold higher (app. k_{cat} of 28 min^{-1} - 140 min^{-1}) at 2.5 to 7.7 fold higher catalytic efficiency (Table S3).

In contrast, PnC_KS-AT showed pronounced promiscuity and, except for **3f**, hydrolytic activity was detected with all tested extender units, including **2** which is not incorporated into the pentaketide (Table 1). The app. k_{cat} and in particular the app. K_M strongly differed, depending on the tested substrate. Interestingly, the app. k_{cat} for **2**, **3a**-**3e** differed only by a factor of four, ranging between 0.4 min^{-1} to 1.4 min^{-1}, while strong differences were observed in the app. K_M, that varied by a factor of 12 between the different substrates, ranging from 6 µM to 73 µM. Transacylation rates of PnC_KS-AT could not be measured with the SucC/SucD-based assay, due to high background activity, which could not be reduced even by further purification of the enzymes. Thus, we determined the app. k_{cat} for transacylation for **2** and **3b**-**3e** by an HPLC-based assay (Figure S8, Figure S9 and Figure S10). PnC_KS-AT was able to transfer all tested substrates, however at lower rates compared to transacylation rates of PnB_KS-AT1, PnB_KS-AT2 and PnD_KS-AT. The highest transacylation rates of PnC_KS-AT were observed for its natural substrate **3b** (4±0.1 min^{-1}) The slightly longer substrates **3c** and **3d** were transacylated at approximately 7- to 12- fold and **2** and **3e** at approximately 100- to 80- fold lower rates, respectively, suggesting that transacylation (more than hydrolysis) is a crucial factor for incorporation by PnC. We independently confirmed these findings by mass spectrometric analysis of acylated PnC_ACP (Table 1, Table S4).

Overall, our kinetic data show that malonyl-CoA specific domains (PnB_KS-AT1, PnB_KS-AT2, and PnD_KS-AT) exclusively transfer malonate to their cognate ACP and neither show transacylation nor hydrolysis for non-native substrates. PnC_KS-AT on the other hand transfers all tested substrates onto its cognate ACP, including **2** which is not incorporated into the polyketide product. PnC_KS-AT also displays hydrolytic activity against all substrates, indicating that in the natural context selectivity is achieved by the ratio of transacylation and hydrolysis rate in PnC.

Table 1. app. K_M and app. k_{cat} values for hydrolysis and transacylation by PnC_KS-AT

Hydrolysis			
Substrate	**K_M [µM]**	**k_{cat} [min^{-1}]**	**k_{cat}/ K_M**
2	53±41	0.6±0.15	0.01
3a	73±33	1.4±0.3	0.02
3b	6±1.8	0.9±0.06	0.15
3c	6.2±1.4	0.7±0.04	0.11
3d	20±6.7	0.8±0.08	0.04
3e	19±15.5	0.4±0.08	0.02
3f	-[a]	-	-
Transacylation			
Substrate	**K_M [µM]**	**k_{cat} [min^{-1}]**	**k_{cat}/ K_M**
2	n.d[b]	0.04±0.002	-
3a	n.d	n.d	-
3b	n.d	4±0.1	-
3c	n.d	0.6±0.03	-
3d	n.d	0.3±0.07	-
3e	n.d	0.05±0.001	-
3f	n.d	-	-

[a] not detected. [b] not determined

Conclusions

In this study, we successfully reconstituted the first six modules of phoslactomycin PKS as a new *in vitro* model system to assess how different extender units are selectively recruited within one PKS system. Establishing this model system required production of various soluble and functional proteins, as well as their respective substrates. *C*-terminal chimeric fusions of DEBS_TE to PnB, PnC and PnD enabled production of different polyketide products ranging from tetra- to hexaketide (Scheme 2), to study incorporation selectivity of natural and non-native extender units in non-competitive and competitive assays.

Experiments on substrate selectivity show that PnB module 2, module 3 and PnD exclusively accept **2**, even in the presence of alternative extender units, demonstrated by *in vitro* reconstitution and kinetic assays. PnC on the other hand shows a high substrate tolerance towards *α*-substituted malonyl-CoA derivatives but does not accept **2** itself, which is present in the cell. Strikingly PnC mainly excludes substrate **3a**, while it readily accepts its natural substrate **3b** and longer derivatives (**3c**-**e**) (Figure 1A, strategy 1). Competition assays show that PnC prefers ethyl-residues over shorter (methyl-) and very long (hexyl-) residues, while in competition PnC does not have a pronounced selectivity against linear and branched medium chain lengths (butyl- and 3-methylbutyl residues) (Figure 1A, strategy 2, 3). The same competition assays also indicate that PnD readily accepts non-natural polyketide intermediates from PnC_ACP, with a preference of shorter side chains at C-2 position (Figure 1B). The preference of PnD for short side chains at C-2 position could be located at the KS and ketoreductase, as previously hypothesized [8]. Thus, the observed trends in product distribution are the result of medium chain length preference of PnC during extender unit incorporation and selection for short(er) side chains at the C-2 position of the polyketide by PnD during downstream processing.

Detailed kinetic analysis shed light on the question of how PnC discriminates against malonate. Hydrolysis rates of PnC_KS-AT between different substrates differ only by a factor of four. Note that hydrolysis of natural substrate **3b** is even two-fold higher than for **2**. Assuming saturating levels of malonyl-CoA on PnC_KS-AT this suggests that substrate preference is guided by differences in transacylation rates of **3b** and **2**. Transacylation of the natural substrate **3b** is 100-fold higher than **2**. The relative

transacylation to hydrolysis rate still shows a 63-fold preferred reaction for **3b** (4.4) versus **2** (0.07). The relative transacylation to hydrolysis rate also explains incorporation of the other substrates in following order **3c** (0.82), **3d** (0.41) and **3e** (0.142) that we observed in the pentaketide system. While the hydrolytic side reaction is often referred to as a proofreading function[7], it is apparently not the hydrolysis per se in PnC_KS-AT that discriminates against a substrate, but rather the efficiency of transacylation. Apparently, PnC_KS-AT was selected against transacylation of malonyl-CoA, which is of importance in natural context in respect to formation of the correct product. The observed flexibility in transacylation (and incorporation) of non-native extender units in PnC_AT and its downstream processing is probably due to the fact that these are not present in the natural context.

Substrate selectivity is also linked to unique sequence motifs. A common motif in AT domains is GX_1SX_2G around the active site serine, whereas X_1 mostly is a histidine [17]. Another important motif is a highly conserved HAFH motif in the binding pocket of malonyl-CoA specific domains. Ethylmalonyl-CoA specific ATs show a less well preserved motif at this position, generally XAGH, with X being F, T, V or H [18]. On both positions PnC_AT shows unique sequence motifs, with a GSS and a CASH motif in the acyl-CoA binding pocket (Figure S11). These sequence differences provide strong arguments for the observed substrate tolerance and render PnC_AT a highly interesting enzyme for further functional investigations.

In sum, we established an additional, complementary *in vitro* PKS system for the detailed study of substrate specificity. This system will hopefully enable the testing of existing hypotheses on extender substrate selectivity and facilitate PKS engineering by offering the possibility of site specific incorporation of alternative extender units to create novel polyketide structures in the future.

Acknowledgements

We thank P. D. Gerlinger, and B. Vögeli for contributing in extensive and fruitful discussions. We also want to thank Dr. N. S. Cortina for running the LC-MS samples and Dr. U. Linne for the mass spectrometric measurement of the ACP domains. This work was supported by the Max Planck Society and the LOEWE program (Landes-Offensive

zur Entwicklung wissenschaftlich-ökonomischer Exzellenz) of the state of Hessen within the framework of the MegaSyn Research Cluster.

References

1. Staunton, J.; Weissman, K. J., Polyketide biosynthesis: a millennium review. *Natural product reports* **2001,** *18* (4), 380-416.
2. Hertweck, C., The biosynthetic logic of polyketide diversity. *Angew Chem Int Ed Engl* **2009,** *48* (26), 4688-716.
3. Oliynyk, M.; Stark, C. B.; Bhatt, A.; Jones, M. A.; Hughes-Thomas, Z. A.; Wilkinson, C.; Oliynyk, Z.; Demydchuk, Y.; Staunton, J.; Leadlay, P. F., Analysis of the biosynthetic gene cluster for the polyether antibiotic monensin in Streptomyces cinnamonensis and evidence for the role of monB and monC genes in oxidative cyclization. *Molecular microbiology* **2003,** *49* (5), 1179-90.
4. Ismail-Ali, A.; Fansa, E. K.; Pryk, N.; Yahiaoui, S.; Kushnir, S.; Pflieger, M.; Wittinghofer, A.; Schulz, F., Biosynthesis-driven structure-activity relationship study of premonensin-derivatives. *Org Biomol Chem* **2016,** *14* (32), 7671-5.
5. Lowry, B.; Robbins, T.; Weng, C. H.; O'Brien, R. V.; Cane, D. E.; Khosla, C., In vitro reconstitution and analysis of the 6-deoxyerythronolide B synthase. *J Am Chem Soc* **2013,** *135* (45), 16809-12.
6. Koryakina, I.; McArthur, J. B.; Draelos, M. M.; Williams, G. J., Promiscuity of a modular polyketide synthase towards natural and non-natural extender units. *Org Biomol Chem* **2013,** *11* (27), 4449-58.
7. Bonnett, S. A.; Rath, C. M.; Shareef, A. R.; Joels, J. R.; Chemler, J. A.; Hakansson, K.; Reynolds, K.; Sherman, D. H., Acyl-CoA subunit selectivity in the pikromycin polyketide synthase PikAIV: steady-state kinetics and active-site occupancy analysis by FTICR-MS. *Chemistry & biology* **2011,** *18* (9), 1075-81.
8. Kalkreuter, E.; CroweTipton, J. M.; Lowell, A. N., Engineering the Substrate Specificity of a Modular Polyketide Synthase for Installation of Consecutive Non-Natural Extender Units. **2019,** *141* (5), 1961-1969.
9. Fushimi, S.; Furihata, K.; Seto, H., Studies on new phosphate ester antifungal antibiotics phoslactomycins. II. Structure elucidation of phoslactomycins A to F. *The Journal of antibiotics* **1989,** *42* (7), 1026-36.
10. Chen, Y. L.; Zhao, J.; Liu, W.; Gao, J. F.; Tao, L. M.; Pan, H. X.; Tang, G. L., Identification of phoslactomycin biosynthetic gene clusters from Streptomyces platensis SAM-0654 and characterization of PnR1 and PnR2 as positive transcriptional regulators. *Gene* **2012,** *509* (2), 195-200.
11. Pfeifer, B. A.; Admiraal, S. J.; Gramajo, H.; Cane, D. E.; Khosla, C., Biosynthesis of complex polyketides in a metabolically engineered strain of E. coli. *Science* **2001,** *291* (5509), 1790-2.
12. Dunn, B. J.; Watts, K. R.; Robbins, T.; Cane, D. E.; Khosla, C., Comparative analysis of the substrate specificity of trans- versus cis-acyltransferases of assembly line polyketide synthases. *Biochemistry* **2014,** *53* (23), 3796-806.
13. Weissman, K. J.; Smith, C. J.; Hanefeld, U.; Aggarwal, R.; Bycroft, M.; Staunton, J.; Leadlay, P. F., The Thioesterase of the Erythromycin-Producing Polyketide Synthase:

Influence of Acyl Chain Structure on the Mode of Release of Substrate Analogues from the Acyl Enzyme Intermediates. *Angew Chem Int Ed Engl* **1998,** *37* (10), 1437-1440.
14. Martin, C. J.; Timoney, M. C.; Sheridan, R. M.; Kendrew, S. G.; Wilkinson, B.; Staunton, J. C.; Leadlay, P. F., Heterologous expression in Saccharopolyspora erythraea of a pentaketide synthase derived from the spinosyn polyketide synthase. *Org Biomol Chem* **2003,** *1* (23), 4144-7.
15. Rittner, A.; Paithankar, K. S.; Huu, K. V.; Grininger, M., Characterization of the Polyspecific Transferase of Murine Type I Fatty Acid Synthase (FAS) and Implications for Polyketide Synthase (PKS) Engineering. **2018,** *13* (3), 723-732.
16. Molnos, J.; Gardiner, R.; Dale, G. E.; Lange, R., A continuous coupled enzyme assay for bacterial malonyl-CoA:acyl carrier protein transacylase (FabD). *Anal Biochem* **2003,** *319* (1), 171-6.
17. Reeves, C. D.; Murli, S.; Ashley, G. W.; Piagentini, M.; Hutchinson, C. R.; McDaniel, R., Alteration of the substrate specificity of a modular polyketide synthase acyltransferase domain through site-specific mutations. *Biochemistry* **2001,** *40* (51), 15464-70.
18. Del Vecchio, F.; Petkovic, H.; Kendrew, S. G.; Low, L.; Wilkinson, B.; Lill, R.; Cortes, J.; Rudd, B. A.; Staunton, J.; Leadlay, P. F., Active-site residue, domain and module swaps in modular polyketide synthases. *J Ind Microbiol Biotechnol* **2003,** *30* (8), 489-94.

Supplementary Information

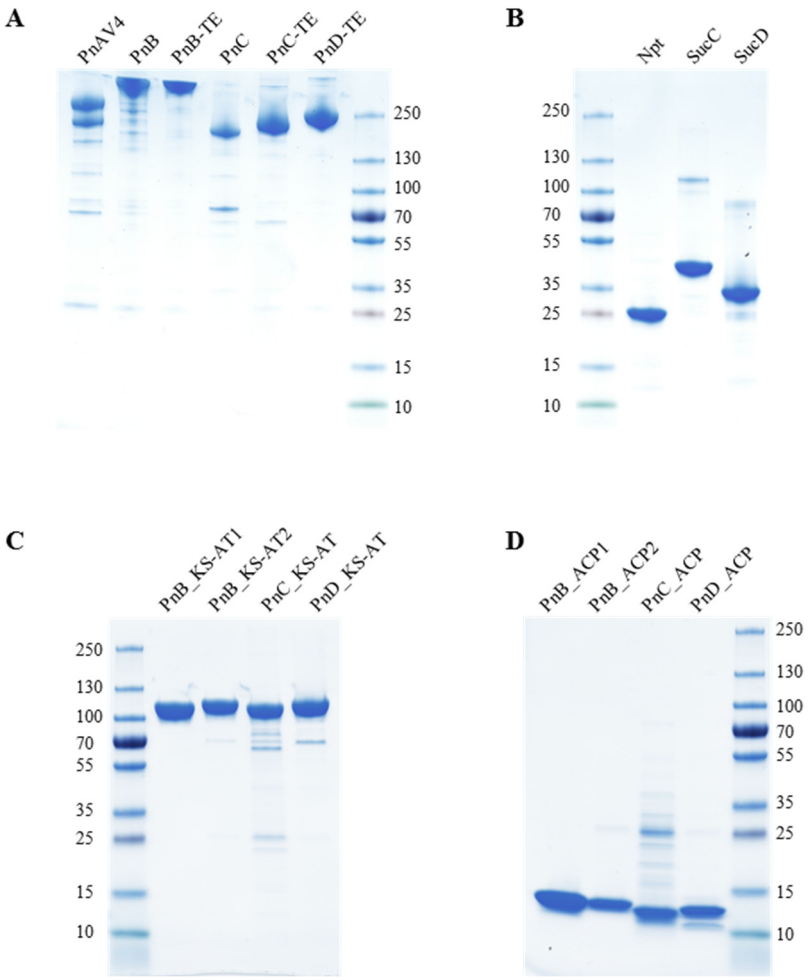


Figure S1: SDS gel of enzymes used in this study. (A) $PnAV_4$ (251.75 kDa), PnB (354.42 kDa), PnB-TE (374.95 kDa), PnC (176.81 kDa), PnC-TE (198.99 kDa) and PnD-TE (221.18 kDa) after affinity purification and size exclusion. (B) Npt (25.78 kDa) after affinity purification, SucC (43.56 kDa) and SucD (31.94 kDa) after affinity purification and size exclusion. (C) PnB_KS-AT1 (97.19 kDa), PnB_KS-AT2 (96.76 kDa), PnC_KS-AT (101.32 kDa), PnD_KS-AT (100.57 kDa) and (D) PnB_ACP1 (12.77 kDa), PnB_ACP2 (12.77 kDa), PnC_ACP (12.01 kDa) and PnD ACP (13.84 kDa) after affinity purification and size exclusion.

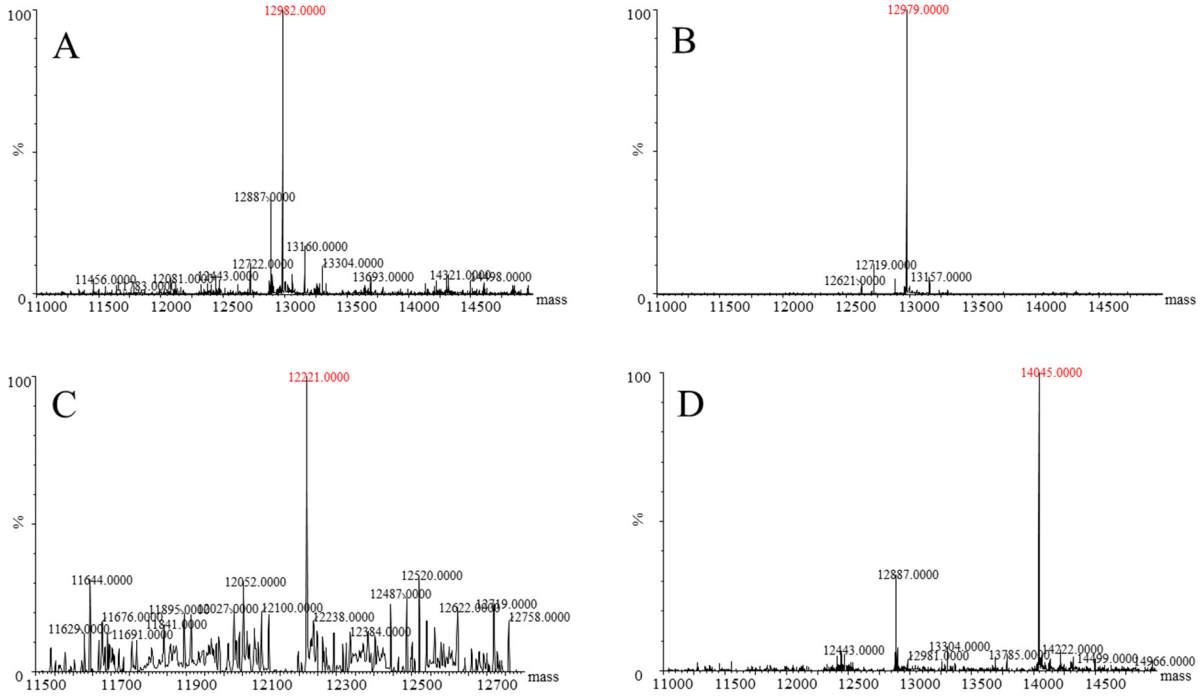


Figure S2: Full conversion of Pn_ACPs to *holo*-ACPs. Npt from *Streptomyces platensis* is able to fully convert *apo*-Pn_ACPs to *holo*-Pn_ACPs. (A) PnB_ACP1, (B) PnB_ACP2, (C) PnC_ACP, (D) PnD_ACP. For no reaction residual *apo*-Pn_ACPs could be detected. 100 µM ACP was incubated with 300 µM CoA, 1 µM Npt in 200 mM NaH_2PO_4 for 30 min at 28°C.

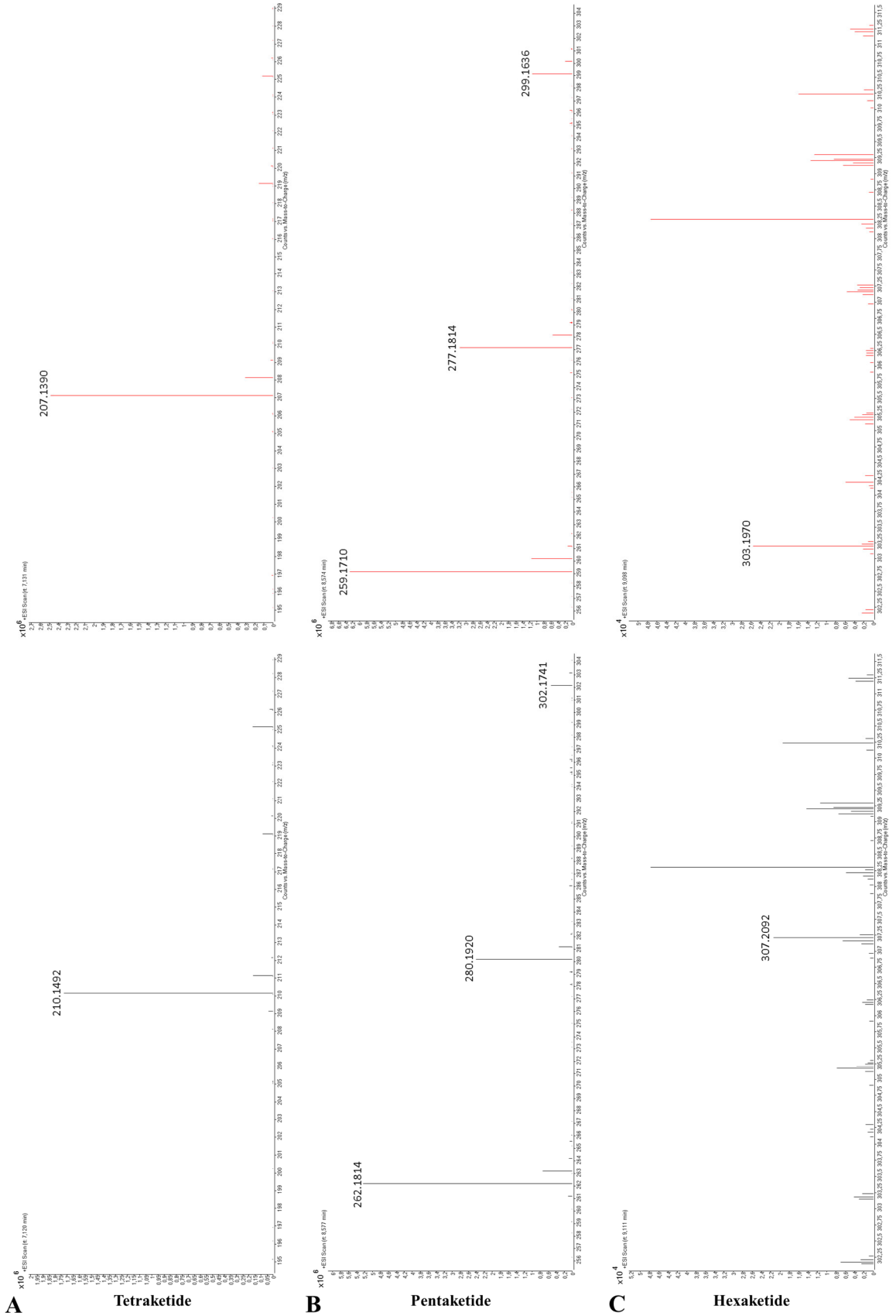

$^{13}C_3$ labelled malonyl-CoA
malonyl-CoA
A
Tetraketide
210.1492
207.1390
B
Pentaketide
262.1814
280.1920
302.1741
259.1710
277.1814
299.1636
C
Hexaketide
307.2092
303.1970

Figure S3: **Extracted ion chromatograms** from assays with PnA_{V4} and PnB (tetraketide system), PnA_{V4}, PnB, PnC-TE_{DEBS} (pentaketide system) and PnA_{V4}, PnB, PnC and PnD-TE_{DEBS} (hexaketide system). Products of the tetraketide and pentaketide system are three times labelled, upon incorporation of 2-C^{13} labelled malonyl-CoA, products of the hexaketide system four times. The tetraketide product (**4**) can only be found as in the dehydrated form (*m/z* $[M+H]^+$ 207.1380). Three masses are shown for the pentaketide product (*m/z* $[M+H]^+$ 277.1798, $[M+Na]^+$ 299.1618 and as *m/z* $[M+H]^+$ 259.1693) (**5b.2**, **5b.4**, Figure S6). For the hexaketide one mass corresponding to the oxidized, dehydrated (**6b.5**, Figure S6) ion is shown as an example (*m/z* $[M+H]^+$ 303.1955).

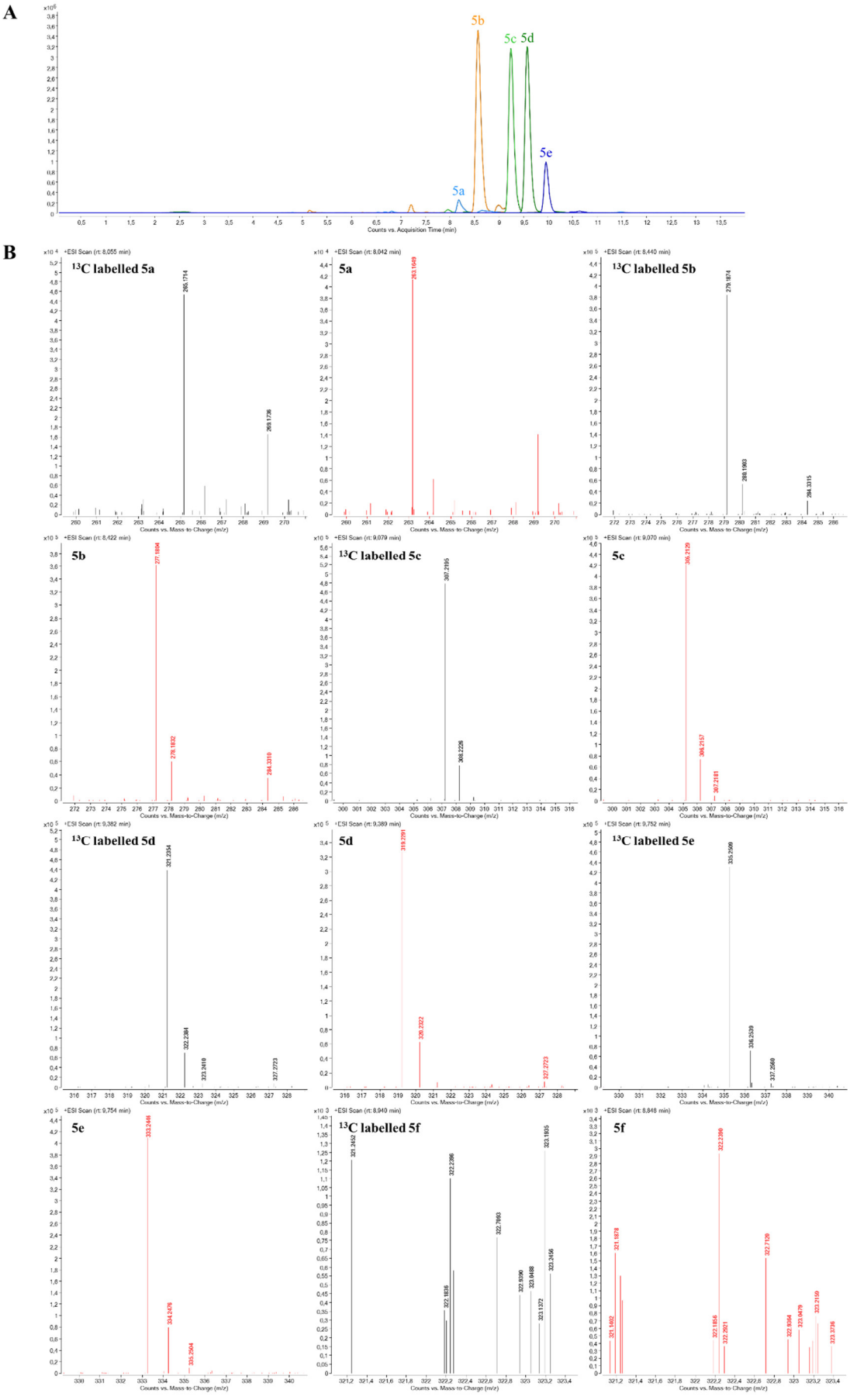

A
5a
5b
5c 5d
5e
Counts vs. Acquisition Time (min)
B
13C labelled 5a
5a
13C labelled 5b
5b
13C labelled 5c
5c
13C labelled 5d
5d
13C labelled 5e
5e
13C labelled 5f
5f
Counts vs. Mass-to-Charge (m/z)

Figure S4: **Mass spectrometric data of pentaketide production.** (A) Overlaid extracted ion chromatograms of **5a**, **5b**, **5c**, **5d** and **5e** from single competition assays. (B) All assays were run with either malonyl-CoA (red) or 2-C^{13} malonyl-CoA (black). This resulted in double labeled pentaketide (PnB module 1 and module 2 incorporate malonyl-CoA). Searched for and shown are the masses of **5a-e.2** and **5f.4** (Figure S6). (**5a**: *m/z* 263.1645 and *m/z* 265.1702, **5b**: *m/z* 277.1798 and *m/z* 279.1859, **5c**: *m/z* 305.2111 and *m/z* 307.2174, **5d**: *m/z* 319.2268 and *m/z* 321.2331, **5e**: *m/z* 333.2424 and *m/z* 335.2487, **5f**: *m/z* 321.1849 and *m/z* 323.1914)

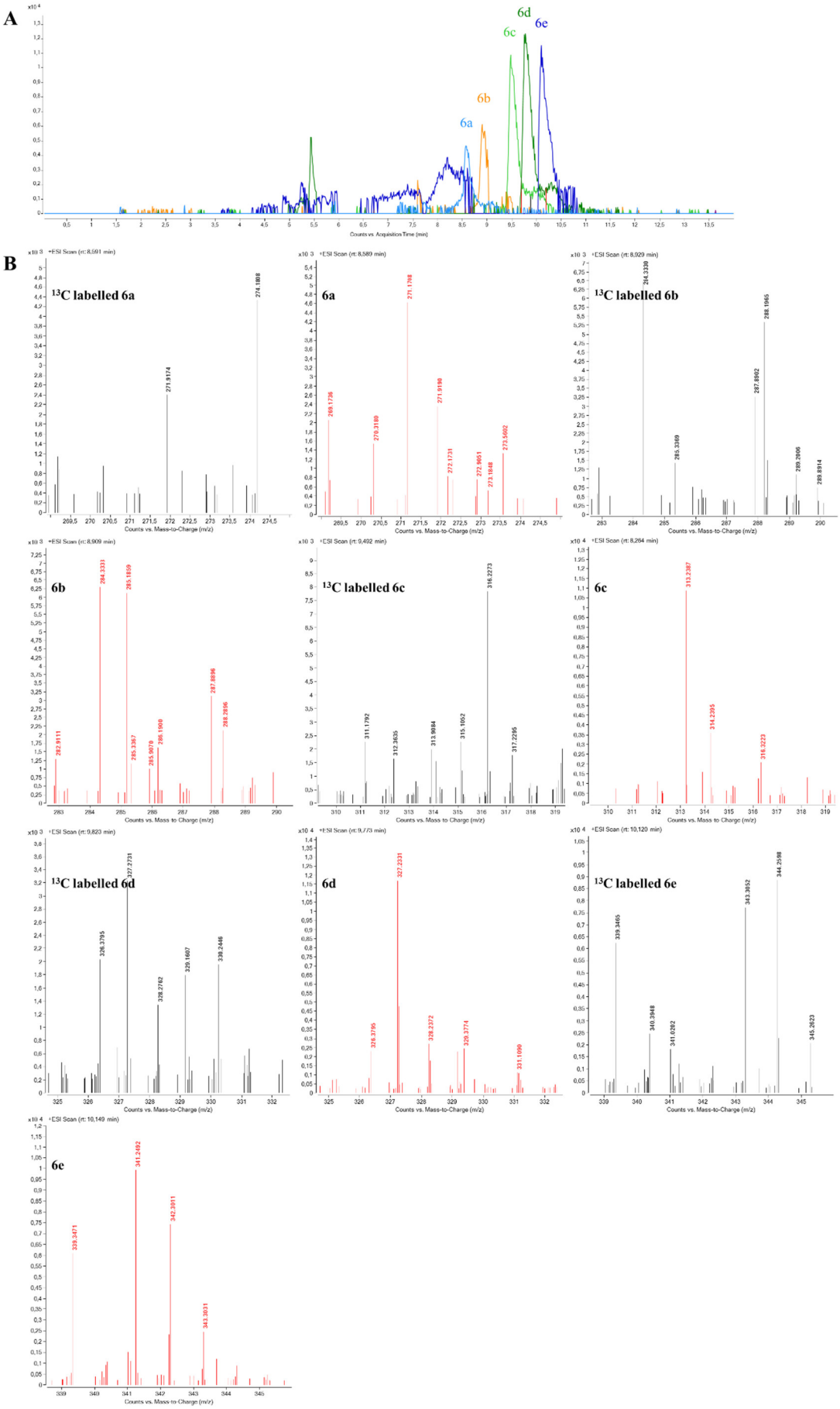
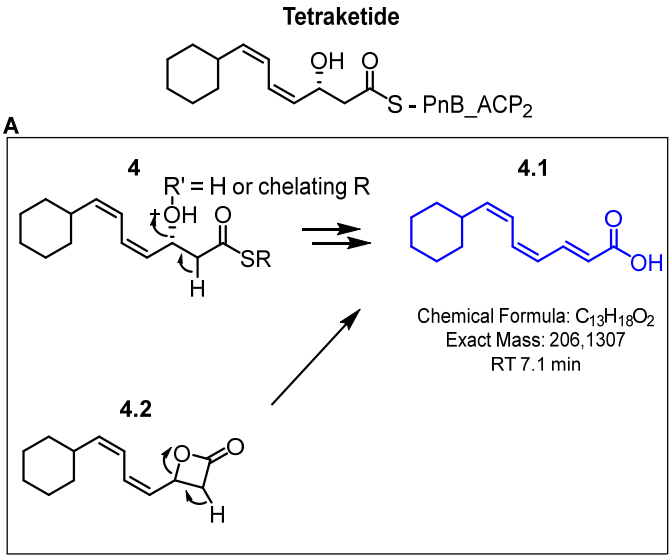
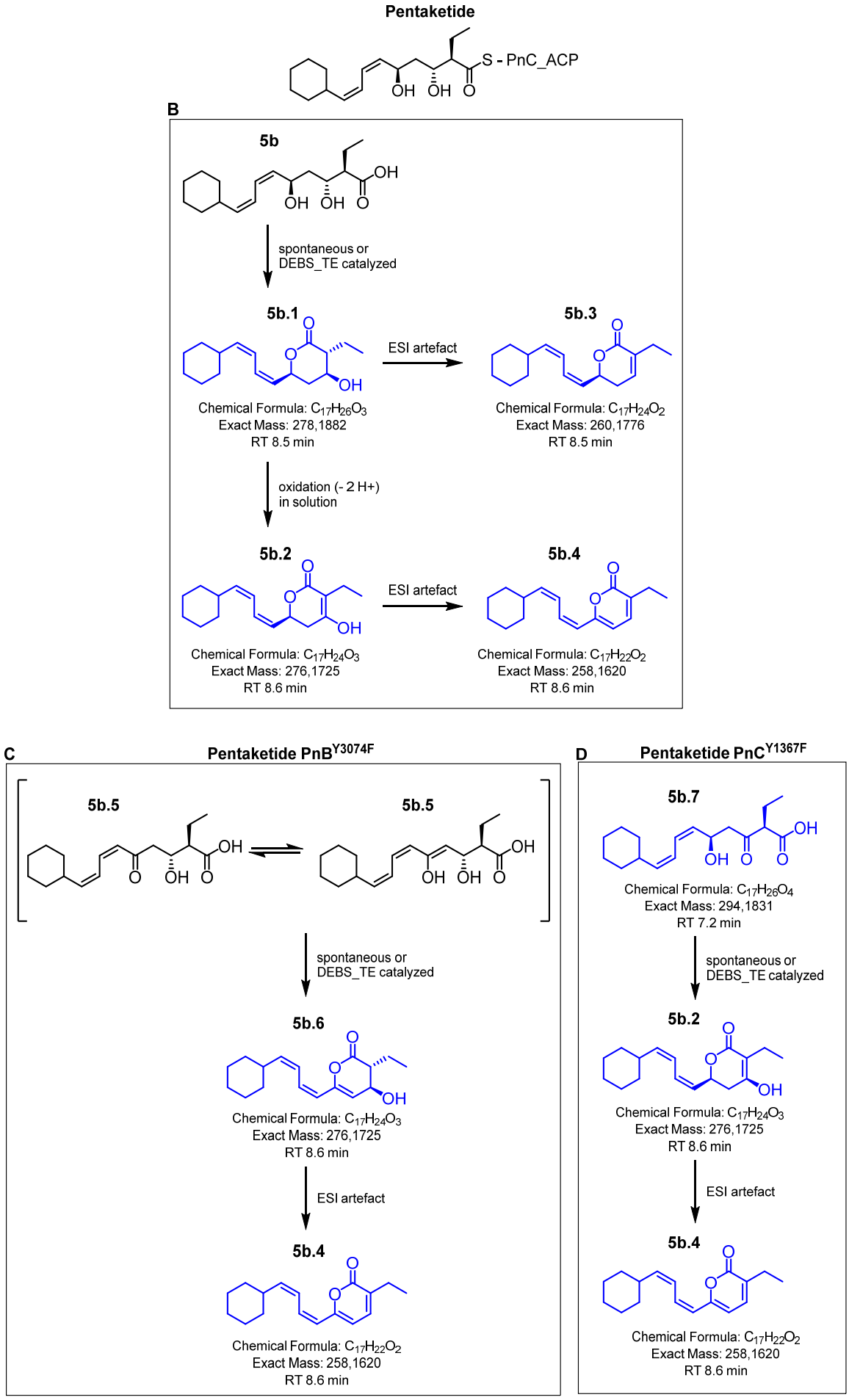

A
6d
6c
6e
6b
6a
Counts vs. Acquisition Time (min)
B
13C labelled 6a
6a
13C labelled 6b
6b
6c
13C labelled 6c
6c
13C labelled 6d
6d
13C labelled 6e
6e
Counts vs. Mass-to-Charge (m/z)

Figure S5: Mass spectrometric data of hexaketide production. (A) Overlaid extracted ion chromatograms of **6a**, **6b**, **6c**, **6d** and **6e** from single extender unit assays. (B) All assays were either run with malonyl-CoA (red) or 2-C^{13} malonyl-CoA (black). This resulted in three time labeled hexaketide (PnB module1 and module 2 and PnD incorporate malonyl-CoA). Searched for and shown are the masses of **6a-e.6** (Figure S6). (**6a**: *m/z* 271.1693 and *m/z* 274.1784, **6b**: *m/z* 285.1849 and *m/z* 288.1941, **6c**: *m/z* 313.2162 and *m/z* 316.2255, **6d**: *m/z* 327.2319 and *m/z* 330.2414, **6e**: *m/z* 341.2475 and *m/z* 344.2569)


Tetraketide
OH
O
S - PnB_ACP2
A
4
R' = H or chelating R
OH
O
SR
H
4.1
O
OH
Chemical Formula: C13H18O2
Exact Mass: 206,1307
RT 7.1 min
4.2
O
O
H


Pentaketide
S - PnC_ACP
OH OH O
B
5b
OH
OH OH O
spontaneous or
DEBS_TE catalyzed
5b.1
ESI artefact
5b.3
OH
Chemical Formula: C17H26O3
Exact Mass: 278,1882
RT 8.5 min
Chemical Formula: C17H24O2
Exact Mass: 260,1776
RT 8.5 min
oxidation (- 2 H+)
in solution
5b.2
5b.4
ESI artefact
OH
Chemical Formula: C17H24O3
Exact Mass: 276,1725
RT 8.6 min
Chemical Formula: C17H22O2
Exact Mass: 258,1620
RT 8.6 min
C
Pentaketide PnB Y3074F
5b.5
5b.5
OH
OH OH O
O OH O
spontaneous or
DEBS_TE catalyzed
5b.6
OH
Chemical Formula: C17H24O3
Exact Mass: 276,1725
RT 8.6 min
ESI artefact
5b.4
Chemical Formula: C17H22O2
Exact Mass: 258,1620
RT 8.6 min
D
Pentaketide PnC Y1367F
5b.7
OH
OH O O
Chemical Formula: C17H26O4
Exact Mass: 294,1831
RT 7.2 min
spontaneous or
DEBS_TE catalyzed
5b.2
OH
Chemical Formula: C17H24O3
Exact Mass: 276,1725
RT 8.6 min
ESI artefact
5b.4
Chemical Formula: C17H22O2
Exact Mass: 258,1620
RT 8.6 min

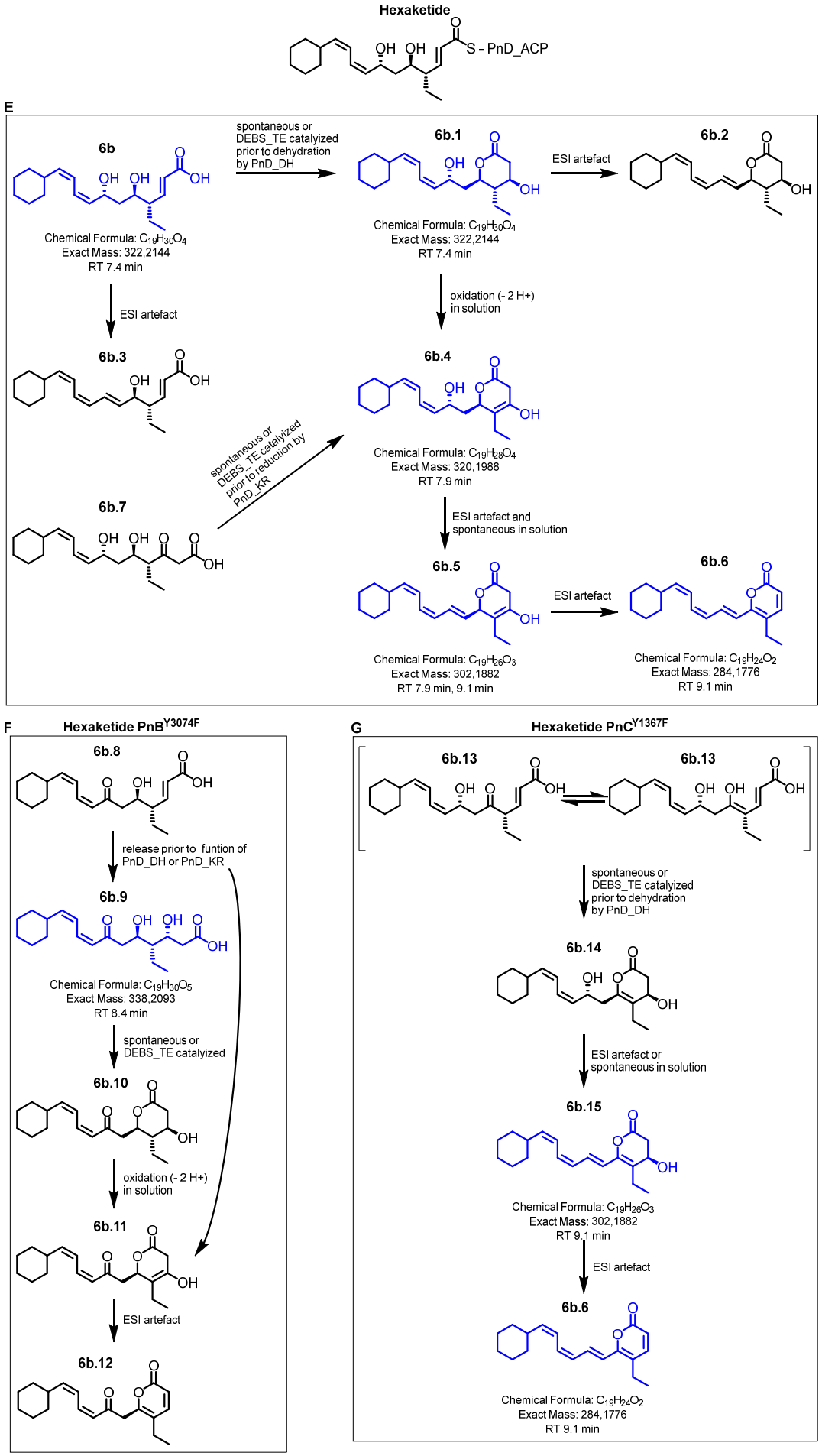

Hexaketide
S - PnD_ACP
E
6b
Chemical Formula: C19H30O4
Exact Mass: 322,2144
RT 7.4 min
spontaneous or DEBS_TE catalyzed prior to dehydration by PnD_DH
6b.1
Chemical Formula: C19H30O4
Exact Mass: 322,2144
RT 7.4 min
ESI artefact
6b.2
ESI artefact
6b.3
oxidation (- 2 H+) in solution
6b.4
Chemical Formula: C19H28O4
Exact Mass: 320,1988
RT 7.9 min
spontaneous or DEBS_TE catalyzed prior to reduction by PnD_KR
6b.7
ESI artefact and spontaneous in solution
6b.5
Chemical Formula: C19H26O3
Exact Mass: 302,1882
RT 7.9 min, 9.1 min
ESI artefact
6b.6
Chemical Formula: C19H24O2
Exact Mass: 284,1776
RT 9.1 min
F
Hexaketide PnB Y3074F
6b.8
release prior to funtion of PnD_DH or PnD_KR
6b.9
Chemical Formula: C19H30O5
Exact Mass: 338,2093
RT 8.4 min
spontaneous or DEBS_TE catalyzed
6b.10
oxidation (- 2 H+) in solution
6b.11
ESI artefact
6b.12
G
Hexaketide PnC Y1367F
6b.13
6b.13
spontaneous or DEBS_TE catalyzed prior to dehydration by PnD_DH
6b.14
ESI artefact or spontaneous in solution
6b.15
Chemical Formula: C19H26O3
Exact Mass: 302,1882
RT 9.1 min
ESI artefact
6b.6
Chemical Formula: C19H24O2
Exact Mass: 284,1776
RT 9.1 min

Figure S6: Product spectrum of phoslactomycin *in vitro* assays. Panel (A), (B) and (E) show the proposed structures of the tetra-, penta- and hexaketide products of the reconstituted phoslactomycin *in vitro* polyketide system. Panel (C), (D), (F) and (G) show the products that are expected when respective KR are catalytically knocked out. Marked in blue, by the formula and retention time (RT), are the masses that were detected in the mass spectrometric measurements, as well as their derived structures. The structures were derived from the proposed biosynthetic pathway for the phoslactomycin polyketide backbone[1].

Tetraketides

(A) After synthesis the tetraketide **4** is bound to the ACP of PnB_ACP2, which can be released as an acid (**4**). Tetraketide **4** can only be found as the dehydrated product **4.1**. Strong conjugation by the double bonds destabilize the hydroxyl group at C-3 and **4.1** is formed. Alternatively, DEBS_TE forms a four membered lactone ring, although such a function was never reported for DEBS_TE.

Pentaketides

(B) The linear pentaketide **5.b** is bound to PnC_ACP after synthesis. However, **5.b** cannot be detected by mass spectrometric analysis, pointing towards lactonization by DEBS_TE and/or spontaneous lactonization, resulting in **5b.1**. Electron spray ionization (ESI) during mass spectrometric measurements causes dehydration, forming **5b.3**, with the same RT as **5b.1**. The hydroxyl group at C-3 of **5b.1** is prone to oxidation in solution, leading to the expected formation of **5b.2**. Further dehydration by ESI leads to formation of **5b.4**, with the same RT as **5b.2**. An alternative route for the synthesis of **5b.2** would be a (partially) inactive PnC_KR. However, since **5b.1** correspond to the reduced pentaketide, it seems rather likely that PnC_KR is fully functional and that **5b.2** is generated by subsequent oxidation of the fully reduced pentaketide.

(C) In the catalytic knock out of the KR of PnB module 2, PnB^{Y3074F}, the C-5 of the pentaketide should carry a ketone group instead of a hydroxyl-group, **5b.5** is bound to PnC_ACP. Keto groups are prone to enolization, enabling the formation of product **5b.6**, either by releasing itself by ring formation from the thioester linkage on the ACP or catalyzed by DEBS_TE. Due to the loss of water by mass spectrometric analysis **5b.4** can also be measured in reaction with PnB^{Y3074F}.

(D) Product **5b.7**, the product of assays with PnC_KR catalytic knockout PnC^{Y1367F}, is bound to PnC_ACP and is the expected product if the free acid is released by DEBS_TE. **5b.7** can be detected in mass spectrometric analysis. This product can form the lactone **5b.2** (as discussed in panel B) by spontaneous lactone ring formation, or catalyzed by DEBS_TE on PnC_ACP. Formation of **5b.4** can be explained by ESI-induced dehydration.

Hexaketides

(E) The linear hexaketide **6b** is bound to PnD_ACP, the acid **6b** and product **6b.1** cannot be distinguished by mass, therefore generation of the double bond by PnD_DH cannot be verified. Product **6b.1** could be a result of spontaneous and/or DEBS_TE catalyzed lactone ring formation prior to dehydration by PnD_DH. The following product **6b.4** could result from **6b.1** by oxidation in solution. An alternative for the formation

of **6b.4** is the release from the thioester on PnD_ACP by ring formation of the unreduced hexaketide **6b.7**, catalyzed by DEBS_TE or spontaneous ring formation. The product **6b.4** can both dehydrate in water (indicated by shift in RT) and by mass spectrometric measurement (same RT), forming **6b.5**. Formation of **6b.6** can be explained by ESI-induced dehydration.

(F) Linear hexaketide **6b.8** with a keto group at C-7 as the product of assays containing PnB^{Y3074F} is bound to PnD_ACP but cannot be found as an acid in solution, however a mass corresponding to **6b.9** can be detected. **6b.9** corresponds to a hexaketide where PnD_DH did not dehydrate the hydroxyl group at C-3, probably due to premature release by DEBS_TE. Following products **6b.10**, **6b.11** and **6c.12** corresponding to lactone ring formation, oxidation or dehydration by ESI (reactions that were observed with the pentaketide) cannot be found. However masses corresponding to product **6b.5** can be found additionally to **6b.9**.

(G) Product **6b.13** as the expected product of assays containing PnC^{Y1367F} would be bound to PnD_ACP but cannot be found by mass spectrometric analysis. Like observed before, the hexaketide could be released by pyrone ring formation of the thioester with the enolized ketogroup at C-5, prior to dehydration by PnD_DH. Subsequently water can be lost to form **6b.15** By ESI measurement another water is lost and **6b.6** is formed (same RT as **5b.15**).

Conclusion to Figure 5: Altogether, our analysis demonstrated that the Pn PKS enzymes are functional *in vitro* and can be combined stepwise to produce tetra- penta- and hexaketide Pn polyketide derivatives. In all cases, we detected dehydrated and oxidized derivatives of the expected polyketides as the main products and are able to propose structures to the calculated molecular formula. To improve formation of the linear hexaketide, we created a catalytic knockout of the PnC_KR (PnC^{Y1367F}). Combining PnC^{Y1367F} with PnD TE_{DEBS} did not improve linear hexaketide formation but led to the formation of **6b.15** (similar spontaneous ring formation is also observed in other PKS [2]). In hexaketide assays, we observed premature release of pentaketide, pointing to autocatalytic lactonization. Lactonization and self-release has also been observed for the DEBS system [3]. To circumvent autocatalytic release and increase hexaketide formation, we decided to remove the hydroxyl group attacking the thioester that is installed by the KR in PnB module 2 (PnB^{Y3074F}). This resulted in production of **6b.9**, albeit total hexaketide production was reduced (18 ± 0.5% residual production). Similar deleterious effects of KR mutants have been observed in other PKS systems [4]. These experiments allowed us to establish and characterize the behavior of the Pn system for subsequent studies on extender unit selectivity.

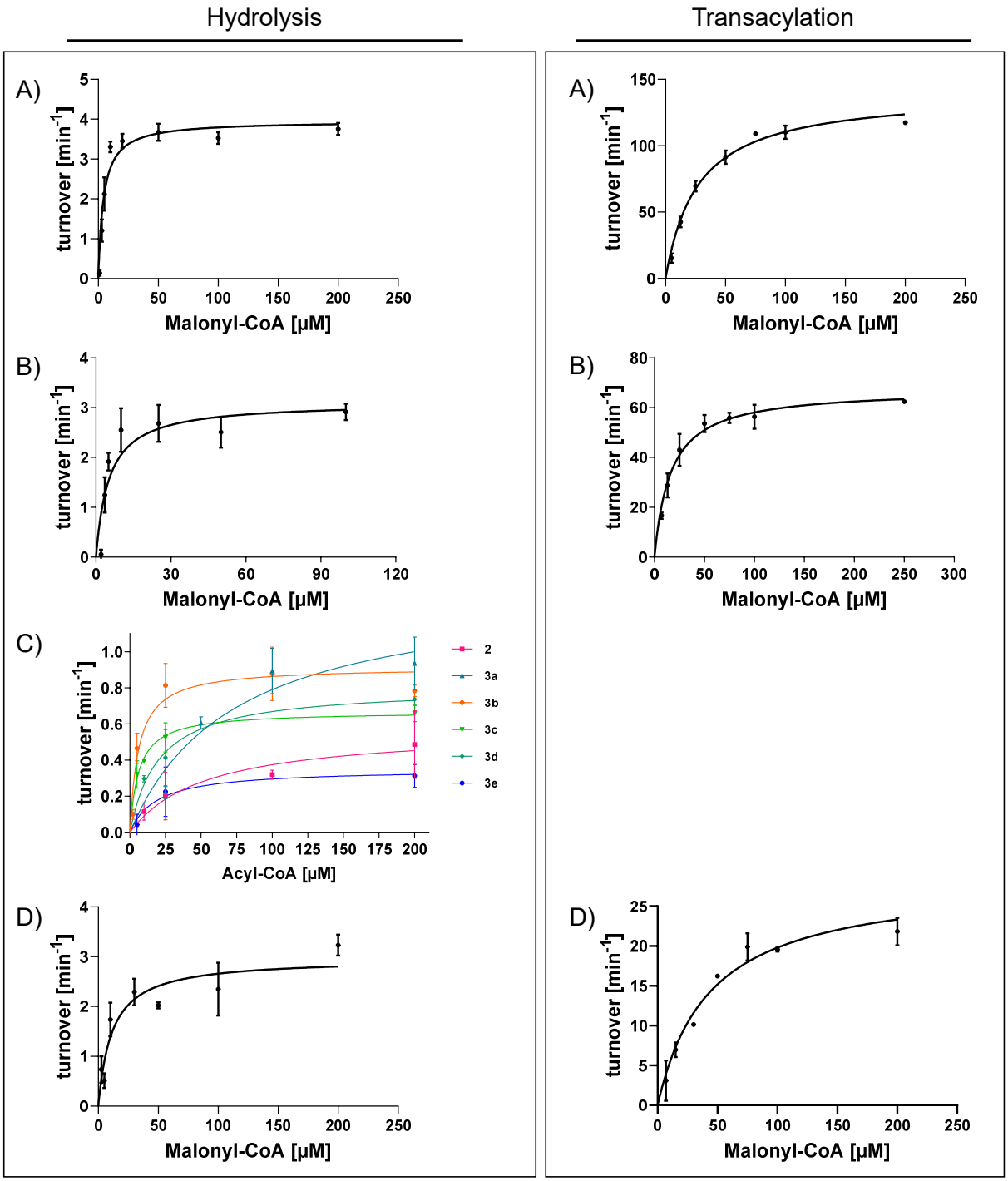


Figure S7: Michaelis Menten plots of KS-AT hydrolysis and transacylation measured with the SucC/SucD based steady state kinetic assay. Each reaction was measured in three technical replicates. (A) PnB_KS-AT1, (B) PnB_KS-AT2, (C), PnC_KS-AT (substrates used are malonyl- (**2**), methylmalonyl- (**3a**), ethylmalonyl- (**3b**), butylmalonyl-CoA (**3c**) 3-methylbutylmalonyl- (**3c**), hexylmalonyl-CoA (**3e)**), (D) PnD_KS-AT

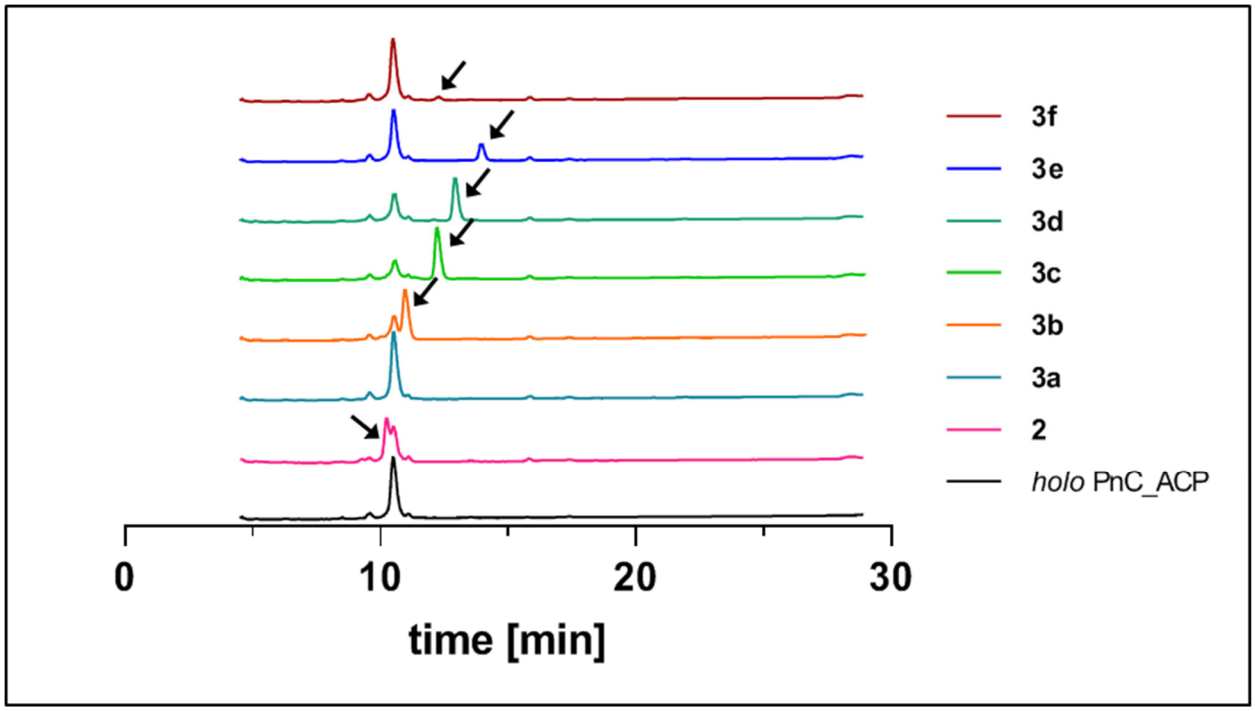


Figure S8: HPLC UV-vis traces of PnC_ACP loaded by PnC_KS-AT (indicated with arrows) with the tested extender units. Incubation time of 5 µM PnC_KS-AT, 100 µM PnC_ACP and 300 µM CoA substrate was 2 h, except for ethylmalonyl-CoA (5 min incubation time) Reaction was run at 25°C. A shift of retention time can be observed for all substrates, except methylmalonyl-PnC_ACP, which elutes at the same retention time as *holo*-PnC_ACP. Substrates used were malonyl- (**2**), methylmalonyl- (**3a**), ethylmalonyl- (**3b**), butylmalonyl-CoA (**3c**) 3-methylbutylmalonyl- (**3c**), hexylmalonyl- (**3e**) and benzylmalonyl-CoA (**3f**).

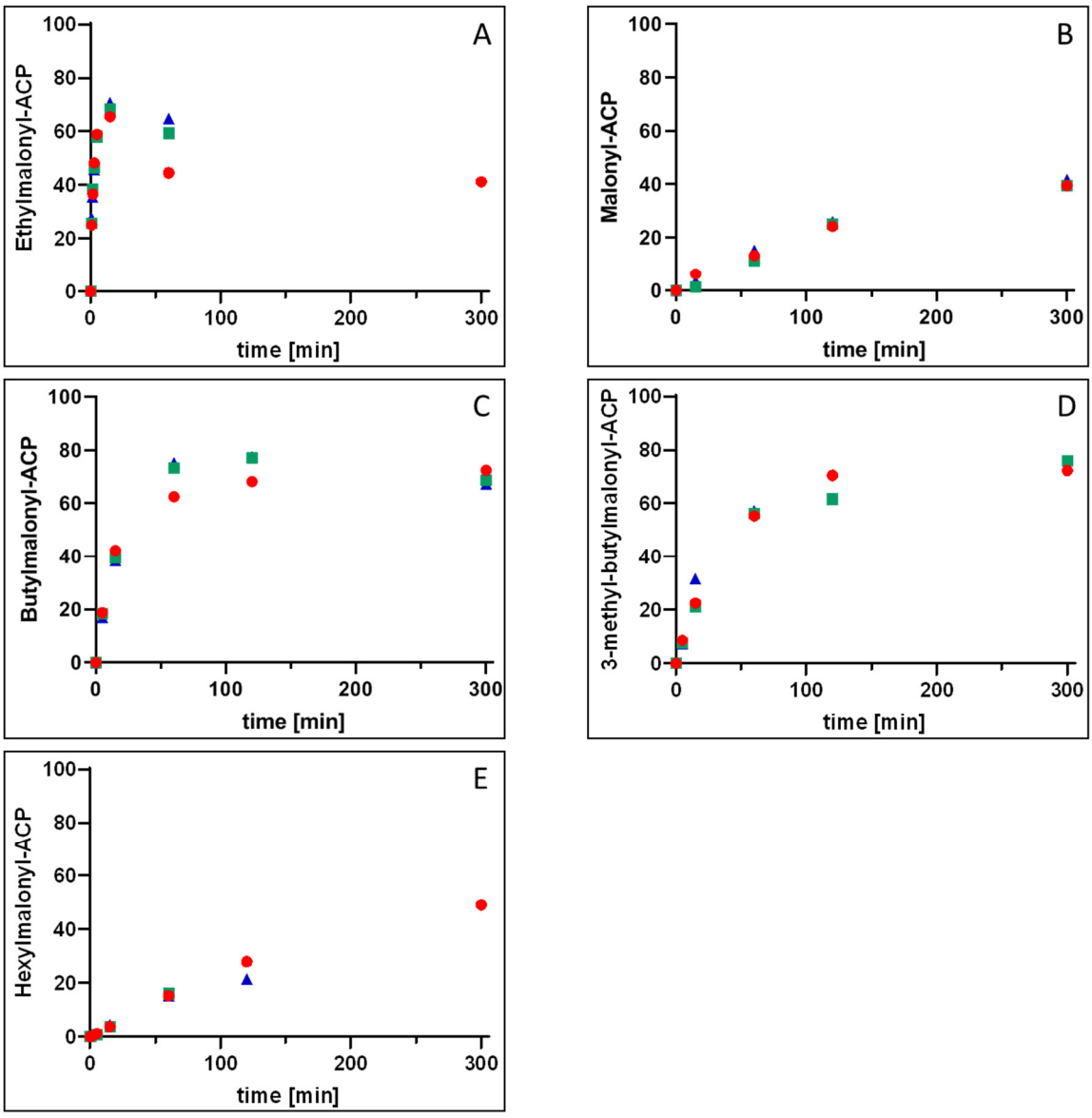


Figure S9: **HPLC UV-vis analysis for PnC_KS-AT transacylation rate**. The three colors represent three technical replicates. Slopes from the linear increase were used for the calculation of k_{cat}. Relative amount of acyl-ACP is given on the y-axis. Substrates used are (A) ethylmalonyl-CoA, (B) malonyl-CoA, (C) butylmalonyl-CoA, (D) 3-methylbutylmalonyl-CoA, (E) hexylmalonyl-CoA. For ethylmalonyl-ACP and butylmalonyl-ACP a decrease of acyl-ACP can be observed after 25 min and after 150 min, respectively. Note also the maximum conversion reached in the transacylation assay varies between approx. 70% for ethylmalonyl-, butylmalonyl-, 3-methylbutylmalonyl-ACP and approx. 40% for malonyl-ACP. Transacylation rate could not be determined for methylmalonyl-CoA, due to same retention time of *holo*-PnC_ACP and methylmalonyl-PnC_ACP (see also Figure S7).

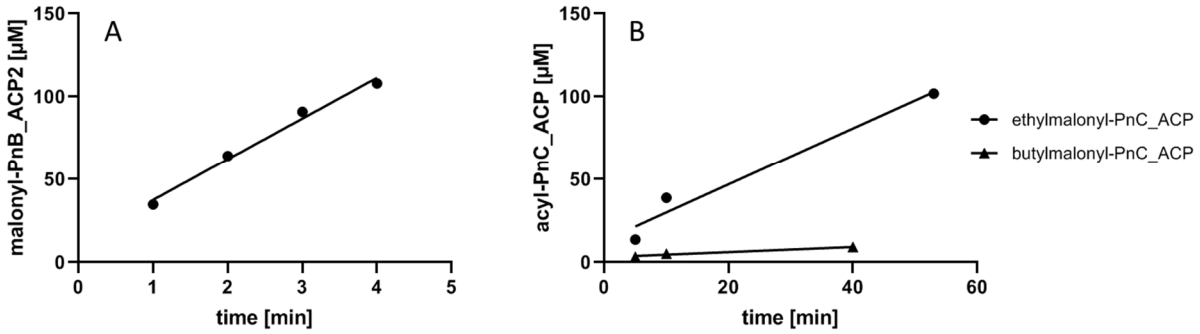


Figure S10: HPLC based transacylation assay to confirm results of SucC/SucD assay. For the measurement of the k_{cat} of transacylation by PnC_KS-AT the SucC/SucD based assay could not be used due to high background activity, which could also not be reduced by repeated size exclusion chromatography of the used proteins. To assure that the results from the SucC/SucD based assay and the HPLC assay are comparable the control described here was done. Turnover of (A) PnB_KS-AT2 (substrate malonyl-CoA) and (B) PnC_KS-AT (substrates used were ethylmalonyl- and butylmalonyl-CoA). 500 nM KS-AT and 400 µM cognate ACP were mixed with 600 µM substrate in 200 mM NaH_2PO_4, pH 7.2 (same protein concentrations in SucC/SucD based assay). The reaction was stopped by adding formic acid and samples were analyzed using HPLC. With the linear slope the turnover was calculated. The k_{cat} of PnB_KS-AT2 with malonyl-CoA was 49.3 per minute and for PnC_KS-AT with ethylmalonyl-CoA 3.4 per minute, with butylmalonyl-CoA 0.31 per minute. With the SucC/SucD based assay a k_{cat} of 74 (PnB_KS-AT2), 4 (PnC_KS-AT with ethylmalonyl-CoA) and 0.6 (PnC_KS-AT with butylmalonyl-CoA) was measured. The determined turnover rates are similar which assures that the measured kinetics displayed in Table S3 and Table 1 can be compared.

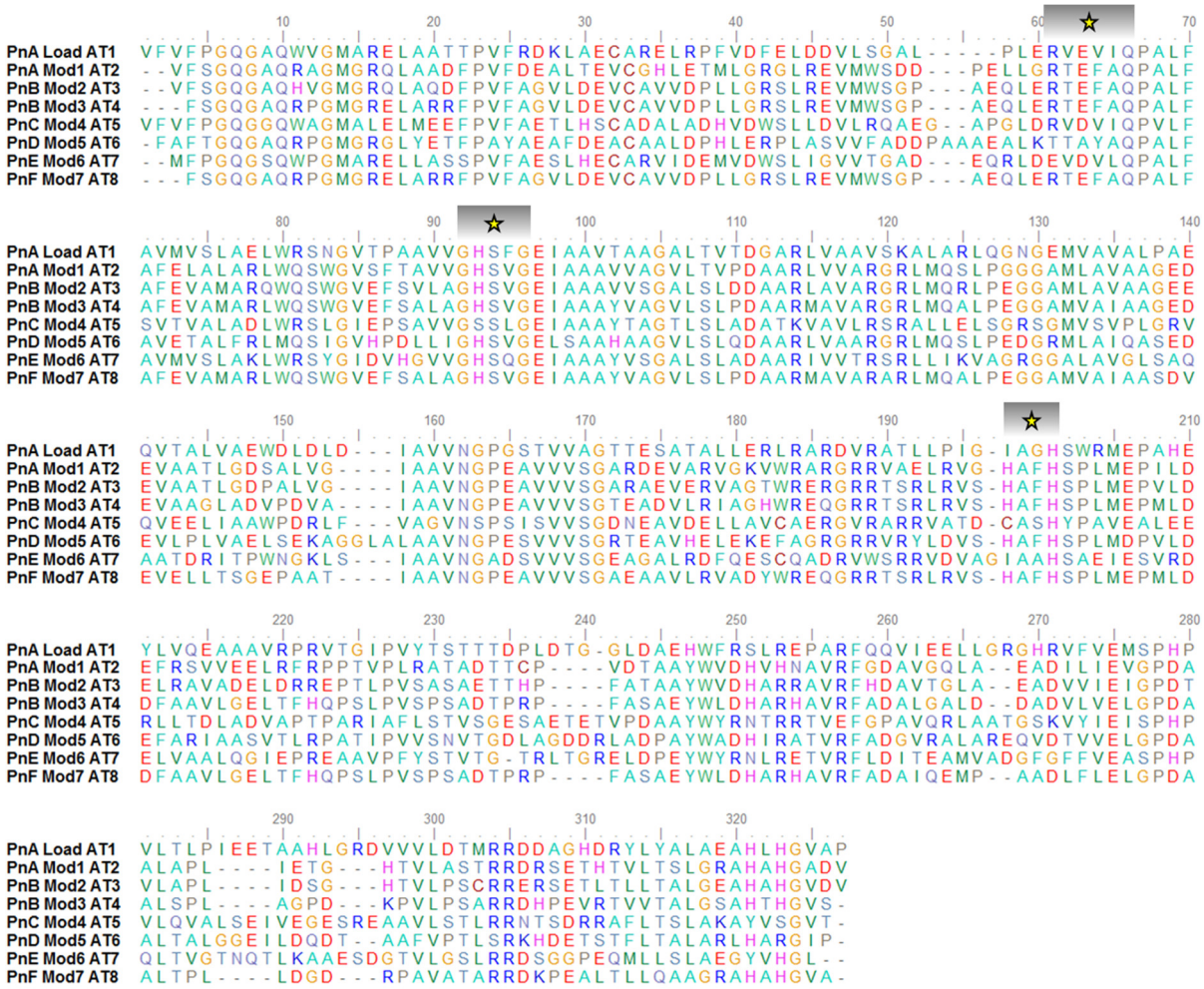
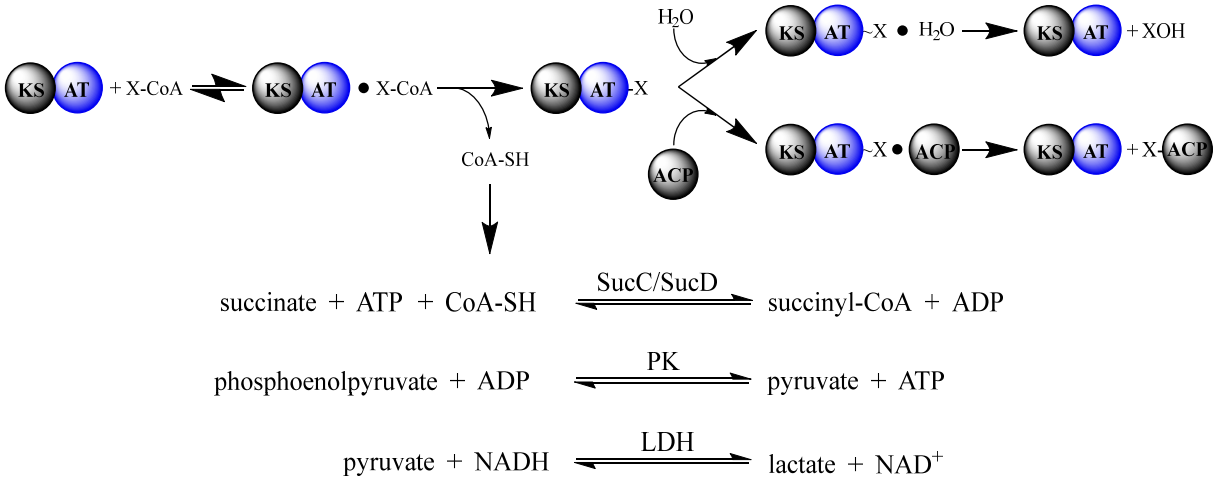


Figure S11: Sequence alignment of AT domains from Pn PKS. Indicated are three extender unit specificity conferring motifs. The canonical GHS can be found in all malonyl-CoA specific Pn_AT domains (amino acid position 92-96), while PnC shows a GSS. Malonyl-CoA specific AT domains contain the highly conserved HAFH motif, methylmalonyl-CoA specific domains most commonly use the YASH motif. Ethylmalonyl-CoA specific AT domains have a less well conserved motif at this position, generally XAGH, with X being F, T, V or H [5]. A motif close to this, IAGH is found in the loading AT domain from PnA. PnC shows the motif CASH in the binding pocket, PnE shows IAAH (amino acid position 198-201).



Scheme S1: Steady state kinetic assay, relying on purified proteins. KS-AT didomains and standalone ACPs are used. Upon covalent binding of the acyl-residue to the AT, CoA-SH is released. With no ACP acceptor available, the acyl-residue is subsequently hydrolyzed by nucleophilic attack of water and the AT can bind a new acyl-residue. If *holo*-ACP is available, the acyl-residue is transferred and the AT can bind a new acyl-residue. Free CoA-SH is converted to succinyl-CoA by *E. coli* SucC/SucD. In this reaction ADP is produced, which can be used by pyruvate kinase/lactic dehydrogenase (PK/LDH) under NADH consumption. For Michaelis Menten plots with data collected with this assay refer to Figure S6.

Table S1: phosphopantetheinyl coverage, selfacylation and transacylation of Pn_ACPs.

	PnB_ACP1	PnB_ACP2	PnC_ACP	PnD_ACP
***holo*-ACP**	26%	75%	96%	66%
Selfacylation of ACP (%)				
M-ACP (2)	0	6.4	8.3	0
MM-ACP (3a)	30	5.7	0	12.5
EM-ACP (3b)	0	0	0	0
BM-ACP (3c)	0	0	0	0
3MBM-ACP (3d)	9.1	0	0	3.5
HM-ACP (3e)	0	0	0	0
BenzM-ACP (3f)	0	0	0	0
Transacylation of ACP (%)				
M-ACP (2)	100	63.8	55.9	19.1
MM-ACP (3a)	35.8	9.3	78.8	6.6
EM-ACP (3b)	0	0	62	0
BM-ACP (3c)	0	0	68.8	0
3MBM-ACP (3d)	9.5	0	60.4	3
HM-ACP (3e)	0	0	16.3	0
BenzM-ACP (3f)	0	0	0	0

Extent of phosphopantetheinlyation of Pn_ACP purified from *E. coli* BAP1. Acyl-ACP buildup upon 2 h incubation of 100 µM *holo*-ACP with 300 µM extender unit at 25°C (“selfacylation”) and acyl-ACP buildup upon 2 h incubation of 5 µM KS-AT, 300 µM CoA thioester, 100 µM ACP at 25°C. The percentage of *apo*- , *holo*- and acyl-ACP was determined by mass spectrometric measurements. Substrates used were malonyl- (**2**), methylmalonyl- (**3a**), ethylmalonyl- (**3b**), butylmalonyl-CoA (**3c**) 3-methylbutylmalonyl- (**3c**), hexylmalonyl- (**3e**) and benzylmalonyl-CoA (**3f**).

Malonyl-, methylmalonyl- and 3-methylmalonyl-ACP can be detected without the presence of an AT domain. This is likely due to transesterification of the acyl-residue from CoA to the thiolgroup of the ACP. In case of malonyl- and methylmalonyl-CoA this is also partly due to sfp mediated loading of the ACP during expression in *E. coli* BAP1. Subtracting already acylated ACPs (“selfacylation”) from the transacylated ACP produced by the AT-mediated transacylation shows that only malonyl-CoA **2** is transferred in the case of PnB_AT1, PnB_AT2 and PnD_AT and that all substrates (BenzM-ACP under detection limit) are transferred in the case of PnC_AT.

Table S2: List of polyketide masses (*m/z*) from the competition assays.

Compound	Retention time [min]	[H+] m/z calculated	[H+] m/z detected	Δ [m/z]	[Na+] m/z calculated	[Na+] m/z detected	Δ [m/z]
4	7.1	207.1380	207.1390	0.0010	229.1199	-	
5b	8.5	279.1955	-		301.1774	301.1757	0.0017
	8.5	261.1849	261.1852	0.0003	283.1669	-	
	8.6	277.1798	277.1802	0.0004	299.1618	299.1624	0.0006
	8.6	259.1693	259.1694	0.0001	281.1512	-	
5a	8.1	265.1798	265.1799	0.0001	287.1618	287.1599	0.0019
	8.1	247.1693	247.1697	0.0004	269.1512	-	
	8.2	263.1642	263.1648	0.0006	285.1461	285.1468	0.0007
	8.2	245.1536	245.1543	0.0007	267.1355	-	
5c	-	307.2268	-		329.2087	-	
	-	289.0313	-		311.1982	-	
	9.2	305.2111	305.2103	0.0008	327.1931	327.1920	0.0011
	9.2	287.2006	287.2001	0.0005	309.1825	-	
5d	-	321.2424	-		343.2243	-	
	-	303.2318	-		325.2138	-	
	9.6	319.2264	319.2269	0.0005	341.2087	341.2079	0.0008
	9.6	301.2162	301.2168	0.0006	323.1981	-	
5e	-	335.2581	-		357.2400	-	
	-	317.2475	-		339.2295	-	
	9.9	333.2424	333.2416	0.0008	355.2244	355.2236	0.0008
	9.9	315.2319	315.2312	0.0007	337.2138	-	
5f	-	341.2111	-		363.1930	-	
	-	323.2005	-		345.1825	-	
	9.0	339.1955	339.1972	0.0017	361.1774	361.1772	0.0002
	9.0	321.1849	321.1857	0.0008	343.1668	-	
6b	7.4	323.2217	-		345.2036	345.2046	0.0010
	7.9	321.2060	-		343.1880	343.1872	0.0008
	7.9; 9.1	303.1955	303.1953	0.0002	325.1774	325.1775	0.0001
	9.1	285.1849	285.1848	0.0001	307.1669	-	
6a	7.1	309.2060	-		331.1880	331.1919	0.0039
	7.6	307.1904	307.1923	0.0019	329.1730	-	
	7.6; 8.7	289.1798	289.1819	0.0021	311.1618	311.1647	0.0029
	8.7	271.1693	271.1718	0.0025	293.1512	-	
6c	-	351.2530	-		373.2349	-	
	8.4	349.2373	-		371.2193	371.2222	0.0029
	9.7	331.2268	331.2220	0.0048	353.2087	353.2109	0.0022
	9.7	313.2162	313.2176	0.0014	335.1982	-	
6d	-	365.2686	-		387.2505		
	8.9	363.2529	-		385.2349	385.2378	0.0029
	10.0	345.2424	345.2441	0.0017	367.2243	367.2281	0.0038
	10.0	327.2319	327.2342	0.0023	349.2138	-	
6e	-	379.2843	-		401.2662	-	
	9.3	377.2686	-		399.2506	-	
	10.1	359.2581	359.2613	0.0032	381.2400	381.2440	0.0040
	10.1	341.2475	341.2530	0.0055	363.2295	-	

Both, *m/z* calculated and detected are given together with the retention time. **4** corresponds to the tetraketide, **5** to the pentaketide and **6** to the hexaketide, **b** indicates ethyl-, **a** methyl-, **c** butyl-, **d** 3-methyl-butyl-, **e** hexyl- and **f** benzyl-residue.

Table S3: full table of app. K_M and app. k_{cat} values for hydrolysis and transacylation by PnB_KS-AT1, PnB_KS-AT2, PnC_KS-AT and PnD_KS-AT.

	Hydrolysis											
	PnB_KS-AT1			PnB_KS-AT2			PnC_KS-AT			PnD_KS-AT		
Substrate	K_M [µM]	k_{cat} [min^{-1}]	k_{cat}/K_M	K_M [µM]	k_{cat} [min^{-1}]	k_{cat}/K_M	K_M [µM]	k_{cat} [min^{-1}]	k_{cat}/K_M	K_M [µM]	k_{cat} [min^{-1}]	k_{cat}/K_M
2	4.3±0.8	4±0.15	0.9	5.2±0.8	3±0.14	0.6	53±41	0.6±0.15	0.01	11.5±3.3	3±0.2	0.26
3a	-[a]	-	-	-	-	-	73±33	1.4±0.3	0.02	-	-	-
3b	-	-	-	-	-	-	6±1.8	0.9±0.06	0.15	-	-	-
3c	-	-	-	-	-	-	6.2±1.4	0.7±0.04	0.11	-	-	-
3d	-	-	-	-	-	-	20±6.7	0.8±0.08	0.04	-	-	-
3e	-	-	-	-	-	-	19±15.5	0.4±0.08	0.02	-	-	-
3f	-	-	-	-	-	-	-	-	-	-	-	-
	Transacylation											
	PnB_KS-AT1			PnB_KS-AT2			PnC_KS-AT			PnD_KS-AT		
Substrate	K_M [µM]	k_{cat} [min^{-1}]	k_{cat}/K_M	K_M [µM]	k_{cat} [min^{-1}]	k_{cat}/K_M	K_M [µM]	k_{cat} [min^{-1}]	k_{cat}/K_M	K_M [µM]	k_{cat} [min^{-1}]	k_{cat}/K_M
2	27±2.6	140±4.3	5.2	16±1.7	74±2	4.6	n.d[b]	0.04±0.002	-	44±7.2	28±1.8	0.64
3a	-	-	-	-	-	-	n.d	n.d	-	-	-	-
3b	-	-	-	-	-	-	n.d	4±0.1	-	-	-	-
3c	-	-	-	-	-	-	n.d	0.6±0.03	-	-	-	-
3d	-	-	-	-	-	-	n.d	0.3±0.07	-	-	-	-
3e	-	-	-	-	-	-	n.d	0.05±0.001	-	-	-	-
3f	-	-	-	-	-	-	-	-	-	-	-	-

[a] not detected. **[b]** not determined

Table S4: Calculated and detected masses of *holo*-, *apo*-, and acyl-ACP.

		PnB_ACP1	PnB_ACP2	PnC_ACP	PnD_ACP
***holo*-ACP**	**calculated**	12983.32	12980.41	12222.47	14046.58
	found	12982.01	12979.0	12221.00	14045.00
***apo*-ACP**	**calculated**	12643.23	12640.32	11882.38	13706.50
	found	12642.00	12639.00	11863.00	13705.00
M-ACP (2)	**calculated**	13069.32	13066.41	12308.47	14132.58
	found	13068.00	13065.00	12307.00	14131.00
MM-ACP (3a)	**calculated**	13083.33	13080.43	12322.48	14146.60
	found	13082.00	13079.00	12321.00	14146.00
EM-ACP (3b)	**calculated**	13097.35	13094.44	12336.50	14160.61
	found	-	-	12335.00	-
BM-ACP (3c)	**calculated**	13125.38	13122.47	12364.53	14188.64
	found	-	-	12364.00	-
3MBM-ACP (3d)	**calculated**	13139.40	13136.49	12378.55	14202.66
	found	13139.00	-	12378.00	-
HM-ACP (3e)	**calculated**	13153.41	13150.50	12392.56	14216.67
	found	-	-	12391.00	-
BenzM-ACP (3f)	**calculated**	13159.37	13156.46	12398.52	14222.63
	found	-	-	12399.00	-

Given are the masses for PnB_ACP1, PnB_ACP2, PnC_ACP and PnD_ACP for the substrates malonyl-CoA **2** (M-ACP), methylmalonyl-CoA **3a** (MM-ACP), ethylmalonyl-CoA **3b** (EM-ACP), butylmalonyl-CoA **3c** (BM-ACP), 3-methylbutylmalonyl-CoA **3d** (3MBM-ACP), hexylmalonyl-CoA **3e** (HM-ACP) and benzylmalonyl-CoA **3f** (BenzM-ACP)

Experimental section

Phoslactomycin PKS *in vitro* assays. The production of phoslactomycin polyketide derivatives was initiated using PnA_{V4} and CHC-CoA (**1**), or PnB and (2*Z*)-cyclohexanepropenyl-SNAC (**7**). The assay was run at a volume of 50 µL and contained the PKS proteins at 9 µM concentration, Npt at 3.5 µM and all substrates at 1 mM concentration (in the case of strategy 3 competition assays, the substrate concentration of *α*-substituted acyl-CoA derivatives was reduced to 0.5 mM to prevent inhibition by high substrate concentrations). The assays additionally contained 5 mM NADPH, 0.1 mM CoA, 5 mM $MgCl_2$ and 100 mM NaH_2PO_4, pH 7.2. The assays were prepared on ice and 10 µL sample was mixed with 1:1 MeOH and stored at -80°C as the negative sample. The assay was run at 25°C overnight, stopped with 1:1 MeOH and stored at -80 °C until analysis. Epimerase, to yield in the production of (2*S*)-acyl-malonyl-CoA derivatives, was

added but did not have an impact on product spectrum and was therefore left out. UPLC-high resolution MS analysis was carried out using an Agilent 6550 iFunnel Q-TOF LC-MS system equipped with an electrospray ionization source set to positive ionization mode. The analyte was separated on a RP-18 column (50 mm x 2.1 mm, particle size 1.7 µm, Kinetex EVO C18, Phenomenex) using a mobile phase system comprised 0.1% formic acid in water (Solvent A) and acetonitrile (Solvent B). Chromatographic separation was carried out using the following gradient condition at a flow rate of 250 µL/min: 0 min 5% B; 1 min 5% B, 6 min 95% B; 6.5 min 95% B; 7 min 5% B. The column oven was set to 40°C and auto sampler was maintained at 8°C. Standard injection volume was 10 µL. Capillary voltage was set at 3.5 kV and nitrogen gas was used as nebulizing (20 psig), drying (13 L/min, 225°C) and sheath gas (12 L/min, 40°C). MS data were acquired with a scan range of 50-1200 *m/z*. LC-MS data were analyzed using MassHunter Qualitative Analysis software (Agilent).

Acyltransferase steady state kinetic assay. The assay buffer (200 mM Tris, glycerol 5%v/v, 20 mM $MgCl_2$, pH 7.4) was used to set up all other solutions for this assay. Four master mixes (MM1 to MM4) in 4x concentration were prepared. For MM1 containing the KS-AT proteins (KS-AT or KS-$AT^{S->A}$ mutants), the enzymes were adjusted to 2 µM (for transacylation) or to 4 µM (for hydrolysis) with assay buffer additionally containing 1 mM DTT and 0.1 mg/mL BSA. For MM2 highly pure and concentrated ACP were adjusted to 1600 µM concentration with the assay buffer. Substrate acyl-CoA containing MM3 was prepared with the assay buffer in varying substrate concentrations ranging from (4x concentration indicated here) 1 µM to 1 mM. MM4 contained 1.6 mM NADH, 20 mM ATP, 30 mM succinate, 10 mM phosphoenolpyruvate (stock solutions should be resuspended in assay buffer, or Tris 200 mM, pH 7.4). First, these compounds were mixed and assay buffer was added, leaving volume for Pyruvate Kinase/Lactic Dehydrogenase (PK/LDH), SucC and SucD. After mixing of the compounds, the enzymes were added to MM4 like following: 0.06 mM SucC, 0.06 mM SucD and 20% v/v of PK/LDH (Sigma-Aldrich) (40 µL PK/LDH for 200 µL MM4). To measure transacylation 25 µL MM3, 25 µL MM4, 25 µL MM2 were mixed and incubated 30 °C for one minute. The reaction was started using 25 µL MM1. To measure hydrolysis 25 µL MM3, 25 µL MM4, 25 µL assay buffer were

mixed, incubated at 30 °C for one minute. The reaction was started with 25 µL MM1. In both cases after starting the reaction, the mixture was immediately transferred into a 1 cm quartz cuvette, preheated to 30 °C and the absorption was monitored at 340 nm with a Cary 60 UV vis (Agilent technologies). The reactions were measured in three to four technical replicates. Simultaneous to the transacylation/hydrolysis reaction a negative control was measured. For transacylation, MM1 was replaced by assay buffer (MM2 caused the highest background activity), for hydrolysis MM3 was replaced by assay buffer. For the transacylation reaction the linear slope during the first minute (seconds 30-60) was used for measurement, for the hydrolysis the linear slope during the first two minutes. The slope of the negative control was subtracted of the slopes measured for the reactions. Prior to measuring the single kinetics of each KS-AT for each substrate, the KS-AT concentration was halved and the velocity was calculated, to assure measurements where the assay is not rate limiting.

HPLC analysis of ACP loading by KS-AT. Loading of *holo*-ACPs was initiated by adding KS-AT with a final concentration of 5 µM to the cognate ACP 100 µM and a single acyl-CoA ester 300 µM. Phosphopantetheinylation of *apo*-ACPs by the 4'-phosphopantetheinyl transferase was initiated by adding Npt with a final concentration of 1 µM to a mix containing 100 µM ACP, 5 mM $MgCl_2$, 200 mM NaH_2PO_4 and 1.5 µM CoA. The reactions were run at 25 °C and stopped by addition of formic acid and injected onto a Discovery BIO Widepore C5 HPLC column. Solvents contained 0.1% TFA. Solvent A was distilled water and Solvent B was acetonitrile. Chromatographic separation was carried out using following gradient condition at a flow rate of 1.5 mL/min: 0 min 5% B; 3 min 35% B; 28 min 50% B; 29 min 95% B; 30 min 5% B. Colum temperature was 40 °C. ACP elution was monitored at 220 nm wavelength.

HPLC kinetic analysis of PnC_KS-AT. The assay contained final concentrations of 100 µM PnC *holo*-ACP, 5 µM PnC KS-AT, 300 µM acyl-CoA thioester and 50 µM NaH_2PO_4, pH 7.2. Reactions were run at 25°C and stopped by addition of formic acid and injected onto a Discovery BIO Widepore C5 HPLC column. Solvents used were distilled water (Solvent A) and acetonitrile (Solvent B), both containing 0.1% TFA.

Chromatographic separation was carried out using following gradient condition at a flow rate of 1.5 mL/min: 0 min 5% B; 3 min 35% B; 28 min 50% B (for butyl-, 3-methylbutyl- and hexylmalonyl-CoA)/ 3 min 36,5% B; 28 min 37,8% B (for malonyl-CoA)/ 3 min 35% B; 28 min 43% B (for ethylmalonyl-CoA). Followed by 29 min 95% B; 30 min 5% B. Colum temperature was 40°C. ACP elution was monitored at 220 nm wavelength. Time points were measured in technical triplicates. The linear increase during the first minutes was used to calculate the transacylation rate.

Mass spectrometric measurements of standalone ACPs. Loading of *holo*- and *apo*-ACPs by KS-AT and Npt was done as described above. All reactions were run at 25°C for 2 h and placed on 4°C until analysis by mass spectrometry. 2 µL of the buffered protein solutions were desalted online using a Waters ACQUITY H-Class HPLC-system equipped with a MassPrep column (Waters). Desalted proteins were eluted into the ESI source of a Synapt G2Si mass spectrometer (Waters) by the following gradient of buffer A (water with 0.05% formic acid) and buffer B (acetonitrile with 0.045% formic acid) at a column temperature of 60°C and a flow rate of 0.1 mL/min: Isocratic elution with 5% A for two minutes, followed by a linear gradient to 95% B within 8 minutes and holding 95% B for additional 4 minutes. Positive ions within the mass range of 500-5000 *m/z* were detected. Glu-Fibrinopeptide B was measured every 45 s for automatic mass drift correction. Averaged spectra were deconvoluted after baseline subtraction and eventually smoothing using MassLynx instrument software with MaxEnt1 extension.

Synthesis of acyl-CoAs. Acyl-CoA synthesis was done like previously described. CHC-CoA was synthesized by chemical CDI coupling of the free acid, analogous as previously explained [6].

Construction of plasmids used in this study. The coding sequence for the phoslactomycin polyketide synthase genes from Streptomyces platensis was synthesized and codon optimized by the Joint Genome Institute. Inserts were amplified by PCR. Oligonucleotides with overhangs for the restriction enzymes NdeI (CATATG) and HindIII (AAGCTT) were used and the agarose gel purified fragments were ligated into the cloning

vector pJET1.2. From there the insert was cut out with NdeI and HindIII and inserted into expression vectors. SucC (Accession number WP_113400154) and SucD (Accession number WP_096861694) coding sequence was amplified by colony-PCR from *E.coli* strain Top10 and cloned following the procedure above. Npt coding sequence was codon optimized for *E.coli* and synthesized. All cloning was done using the *E.coli* strain Top10. Catalytic knock out mutants were constructed using mismatching oligonucleotides, containing the DNA triplet for the desired amino acid. The plasmid was divided into two or three parts and amplified by PCR using Q5® High-Fidelity DNA polymerase. The purified fragments were assembled using the Gibson Assembly ® Master Mix purchased from New England Biolabs.

Vector	Insert	His_6-tag	Protein size (kDA)	Extinction coefficient (ε)
pET21a(+)	PnA_{F4}	C-terminal	251.74	251.79
pET21a(+)	PnB	C-terminal	354.42	401.39
pET21a(+)	PnB^{Y3074F}	C-terminal	354.39	399.90
pET21a(+)	$PnB\text{-}TE_{DEBS}$	C-terminal	374.95	437.96
pET21a(+)	PnC	C-terminal	176.81	202.08
pET21a(+)	PnC^{Y1367F}	C-terminal	176.79	200.59
pET21a(+)	$PnC\text{-}TE_{DEBS}$	C-terminal	198.99	238.65
pET21a(+)	$PnC^{Y1367F}\text{-}TE_{DEBS}$	C-terminal	198.98	237.16
pET21a(+)	PnD	C-terminal	197.11	172.40
pET21a(+)	$PnD\text{-}TE_{DEBS}$	C-terminal	221.18	202.10
pET21a(+)	Npt	C-terminal	25.78	36.11
pET28b(+)	PnB-KSAT1	N-terminal	97.19	102.25
pET28b(+)	PnB^{S669A}-KSAT1	N-terminal	97.17	102.25
pET28b(+)	PnB-ACP1	N-terminal	12.77	/
pET28b(+)	PnB-KSAT2	N-terminal	96.76	89.76
pET28b(+)	PnB^{S671A}-KSAT2	N-terminal	67.74	89.76
pET28b(+)	PnB-ACP2	N-terminal	12.77	/
pET28b(+)	PnC-KSAT	N-terminal	101.32	114.74
pET28b(+)	PnC^{S684A}-KSAT	N-terminal	101.30	114.74
pET28b(+)	PnC-ACP	N-terminal	12.01	/
pET28b(+)	PnD-KSAT	N-terminal	100.57	85.75
pET28b(+)	PnD^{S681A}-KSAT	N-terminal	100.55	85.75
pET28b(+)	PnD-ACP	N-terminal	13.84	/
pET28b(+)	SucC	N-terminal	43.56	27.18
pET28b(+)	SucD	N-terminal	31.94	9.19

Protein expression and purification. Expression constructs encoding for proteins with an ACP domain were transformed into *E.coli* BAP1 or BL21 (DE3) competent cells, all other constructs were transformed into *E.coli* BL21 (DE3) and supplemented with the respective antibiotics. Overnight cultures of 10 mL LB were inoculated with a single colony and used for inoculation of 1 L Terrific Broth medium. Cells were grown at 37 °C until they reached an OD_{600} of 1, induced with 100 µM isopropyl-β-D-thiogalactopyranoside and PKS proteins were expressed at 16°C shaking at 90 rpm for approximately 15 h. Cells were harvested, immediately used for purification or stored at -20°C. The other proteins were expressed at 18°C. SucC and SucD were expressed at 23°C.

Protein Nickel NTA-purification. Cell pellets were resuspended in Buffer A (500 mM NaCl, 50 mM NaH_2PO_4, 10% v/v glycerol, pH 7.5) and lysed by sonication, centrifuged at 42.000 g, 4 °C for 45 min and the supernatant was mixed with 3.5 mL Protino® Ni-NTA Agarose purchased from Macherey Nagel and incubated for 2 h on ice with slow shaking. The beads solution was transferred into Protino® Columns 14 mL, washed with 50 mL Buffer A. Following the beads were washed with 20 mL washing buffer (25 mM Imidazole, 500 mM NaCl, 50 mM NaH_2PO_4, 10% v/v glycerol, pH 7.5). Proteins were eluted with 8 mL Buffer B (500 mM Imidazole, 500 mM NaCl, 50 mM NaH_2PO_4, 10% v/v glycerol, pH 7.5). SucC, SucD, Npt and all standalone ACP proteins were further purified by size exclusion using HiLoad 16/600 Superdex 200 pg column (300 mM NaCl, 25 mM NaH_2PO_4, pH 7.5). All other proteins were subjected to anion exchange. Protein concentration was determined by UV-vis measurements at 280 nm, in the case of ACPs concentration was determined by Bradford assay.

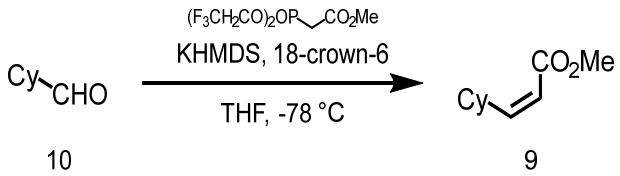
Anion exchange. The eluate from the Nickel beads purification was diluted to 120 mL with Anion A (50 mM NaH_2PO_4, pH 7.5) and loaded onto a 5 mL HiTrap Q HP anion exchange 5 mL chromatography column, purchased from GE Healthcare Life Sciences, with a flow of 3 mL/min. A gradient to 100% Anion B (50 mM NaH_2PO_4, 500 mM NaCl pH 7.5) with a flow of 4 mL/min in 30 min was run and protein containing fractions were collected and concentrated using Amicon® Ultra Centrifugal Filters, purchased from

Merck Millipore. Proteins were used immediately for assays or stored with 30% v/v glycerol in -80 C after shock freezing in liquid nitrogen.

Synthesis of (2*Z*)-cylcohexanepropenyl-SNAC

General Information. All non-aqueous reactions were carried out using flame-dried glassware under argon atmosphere. All solvents were distilled by rotary evaporation. Solvents for non-aqueous reactions were dried as follows prior to use: Tetrahydrofuran (THF) was dried with potassium hydroxide and subsequently distilled from Solvona®. Dichloromethane and triethylamine were distilled from calcium hydride. *N*-acetylcysteamin (HSNAC) was synthesised according to literature [7]. All other commercially obtained reagents were used as received. Reactions were monitored by thin layer chromatography (TLC) using Merck Silica Gel 60 F_{254}-plates and visualised by fluorescence quenching under UV-light. In addition, TLC-plates were stained using a potassium permanganate stain. Chromatographic purification of products was performed on Merck Silica Gel 60 (230-400 mesh) unless otherwise noted using a forced flow of eluents. Concentration *in vacuo* was performed by rotary evaporation at 40 °C and appropriate pressure and by exposing to fine vacuum at room temperature if necessary. NMR spectra were recorded on a Bruker AV 300 MHz*,* AV III 500 MHz, AV III HD 500 MHz spectrometer at room temperature. Chemical shifts are reported in ppm with the solvent resonance as internal standard. Data are reported as follows: s = singlet, d = doublet, t = triplet, q = quartet, quint = quintet, m = multiplet. Mass spectra were recorded by the mass service department of the Philipps-Universität Marburg. HR-ESI mass spectra were acquired with an LTQ-FT mass spectrometer (Thermo Fischer Scientific). The resolution was set to 100 000. IR spectra were recorded on a Bruker IFS 200 spectrometer. The absorption bands are given in wave numbers (cm^{-1}), intensities are reported as follows: s = strong, m = medium, w = weak, br = broad band. Melting points were determined on a Mettler Toledo MP70 using one end closed capillary tubes. Optical rotations were determined at 20 ° with a Krüss P8000-T polarimeter.

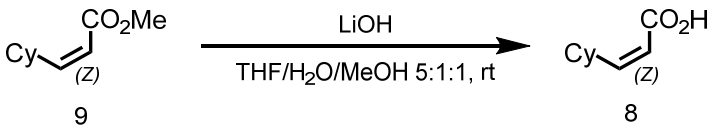
Synthesis of methyl (*Z*)-3-cyclohexylacrylate (9)



10 9

To a solution of Still-Gennari-reagent (0.64 mL, 3.03 mmol, 1.20 eq.) and 18-crown-6 (0.77 g, 2.90 mmol, 1.15 eq.) in tetrahydrofuran (15 mL) at 0 °C under an argon atmosphere was added a solution of potassium bis(trimethylsilyl)amide (0.5 m in toluene, 5.80 mL, 2.90 mmol, 1.15 eq.) dropwise. After 30 min the solution was cooled to -78 °C and aldehyde **10** (0.30 mL, 2.52 mmol, 1.00 eq.) was added dropwise. The solution was stirred 1 h at -78 °C and afterwards reaching room temperature overnight. A sat. solution of ammonium chloride and diethyl ether were added, the layers separated and the aqueous was extracted with diethyl ether (3x). The combined extracts were dried (magnesium sulphate), the volatiles removed *in vacuo* and the residue was purified by flash chromatography (silica gel, pentane/diethyl ether 50:1) to give (*Z*)-ester **9** (0.37 g, 2.20 mmol, 87%) as colourless liquid. TLC: R_f = 0.40 (pent/diethyl ether 50:1); ^{1}H-NMR (500 MHz, $CDCl_3$): δ = 6.03 (dd, *J* = 11.5, 10.0 Hz, 1H, CyC*H*=CH), 5.66 (dd, *J* = 11.5, 1.0 Hz, 1H, CyCH=C*H*), 3.70 (s, 3H, OCH_3), 3.34-3.25 (m, 1H, Cy-C*H*), 1.74-1.64 (m, 5H, Cy-CH_2), 1.40-1.30 (m, 2H, Cy-CH_2), 1.23-1.13 (m, 1H, Cy-CH_2), 1.12-1.03 (m, 2H, Cy-CH_2) ppm; ^{13}C-NMR (125 MHz, $CDCl_3$): δ = 166.8 (*C*O_2Me), 156.1 (Cy*C*H=), 117.1 (CyCH=*C*H), 51.0 (O*C*H_3), 37.3 (Cy-*C*H), 32.3 (2xCy-*C*H_2), 25.9 (Cy-*C*H_2), 25.4 (2xCy-*C*H_2) ppm; HRMS (ESI+): *m/z* calcd. for $C_9H_{13}O$ $[M+Na]^+$ 191.1043, found 191.1042; u_{max}(film): 2965 (w), 2926 (s), 2854 (w), 2358 (m), 2035 (m), 2029 (w), 1728 (s), 1193 (w), 1176 (m), 483 (m), 453 (m), 439 (m), 424 (s) cm^{-1}.

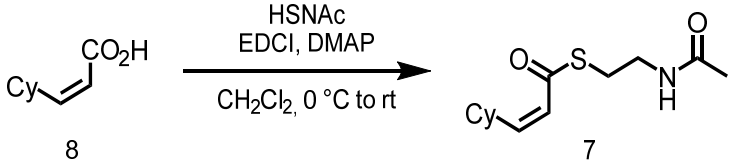
Synthesis of (*Z*)-3-cyclohexylacrylic acid (8*)*



9 8

To a solution of (*Z*)-ester **9** (104 mg, 0.62 mmol, 1.00 eq.) in a 5:1:1 mixture of tetrahydrofuran/water/methanol (7 mL) at room temperature was added lithium hydroxide monohydrate (259 mg, 6.18 mmol, 10.0 eq.) and the mixture was stirred overnight. After addition of water the pH-value was adjusted (pH=3) with hydrochloric acid (1 m), the mixture extracted with diethyl ether (3x) and the combined extracts were dried (magnesium sulphate). The volatiles were removed *in vacuo* and the residue purified by flash chromatography (silica gel, pentane/ethyl acetate 3:1) to give (*Z*)-acid **8** (80 mg, 0.52 mmol, 84%) as white solid. TLC: R_f = 0.55 (pent/ethyl acetate 1:1); ^{1}H-NMR (500 MHz, $CDCl_3$): δ = 6.15 (dd, *J* = 11.3, 10.2 Hz, 1H, Cy-C*H*=), 5.67 (d, *J* = 11.3 Hz, 1H, Cy-CH=C*H*), 3.33-3.26 (m, 1H, Cy-C*H*), 1.73-1.64 (m, 5H, Cy-CH_2), 1.42-1.30 (m, 2H, Cy-CH_2), 1.25-1.06 (m, 3H, Cy-CH_2) ppm (CO_2H was not detectable due to fast exchange with the solvent); ^{13}C-NMR (125 MHz, $CDCl_3$): δ = 170.0 (CO_2H), 158.1 (Cy-*C*H=), 116.6 (Cy-CH=*C*H), 37.4 (Cy-*C*H), 32.3 (2xCy-*C*H_2), 25.9 (Cy-*C*H_2), 25.4 (2xCy-*C*H_2) ppm (*C*O_2H was not detectable due to low C_q-intensity); HRMS (ESI−): *m/z* calcd. for $C_9H_{13}O_2$ $[M]^-$ 153.0921, found 153.0921; υ_{max}(film): 2926 (s), 2852 (w), 2734 (w), 2575 (w), 1694 (s), 1638 (w), 1439 (m), 1351 (w), 1290 (w), 1270 (w), 1242 (m), 1215 (w), 1135 (w), 1029 (w), 965 (w), 931 (w), 892 (w), 827 (w), 798 (w), 729 (w), 695 (w), 451 (w) cm^{-1}; m.p.: 106.5 °C (dichloromethane).

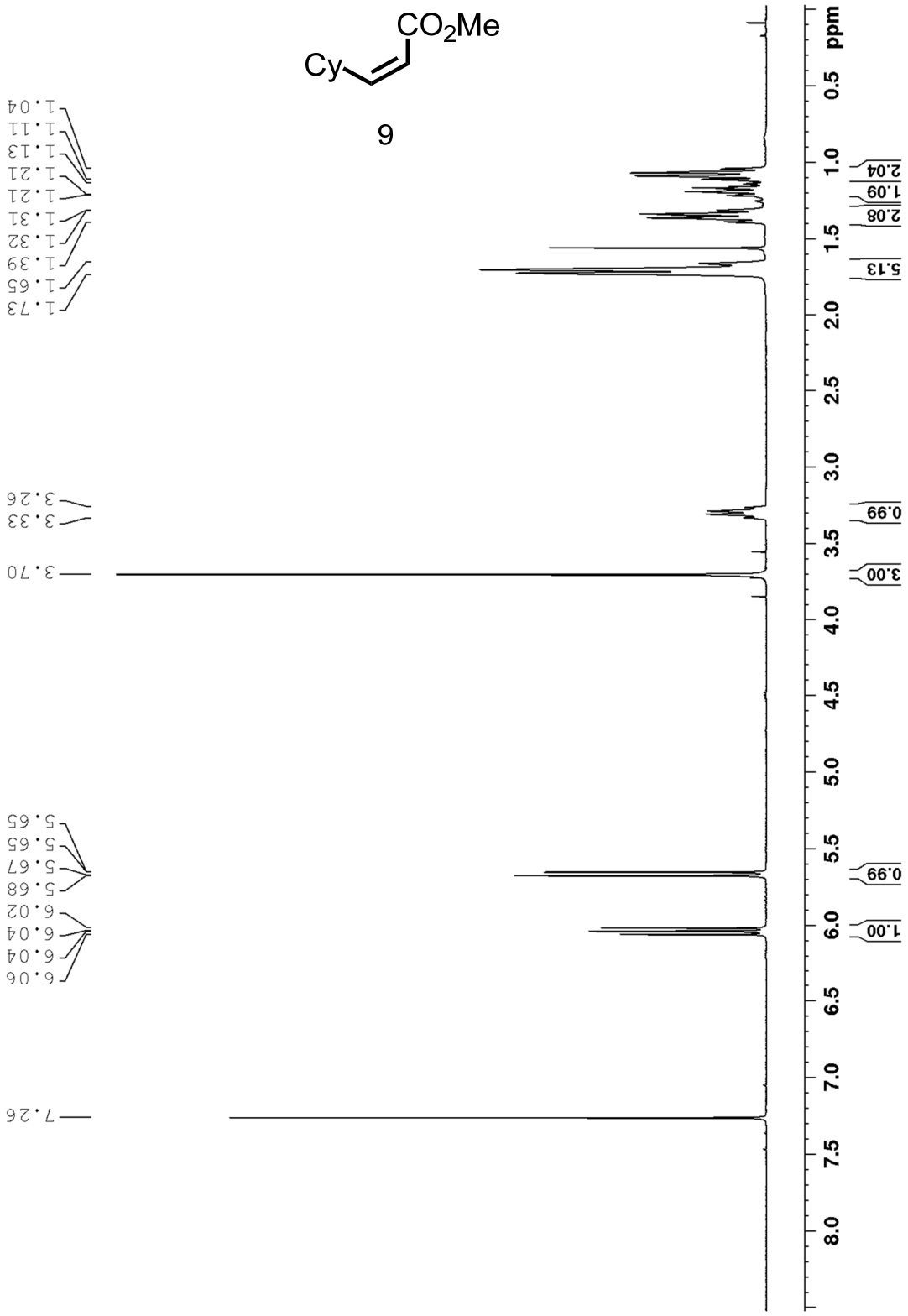
Synthesis of *S*-(2-acetamidoethyl) (*Z*)-3-cyclohexylprop-2-enethioate (7)

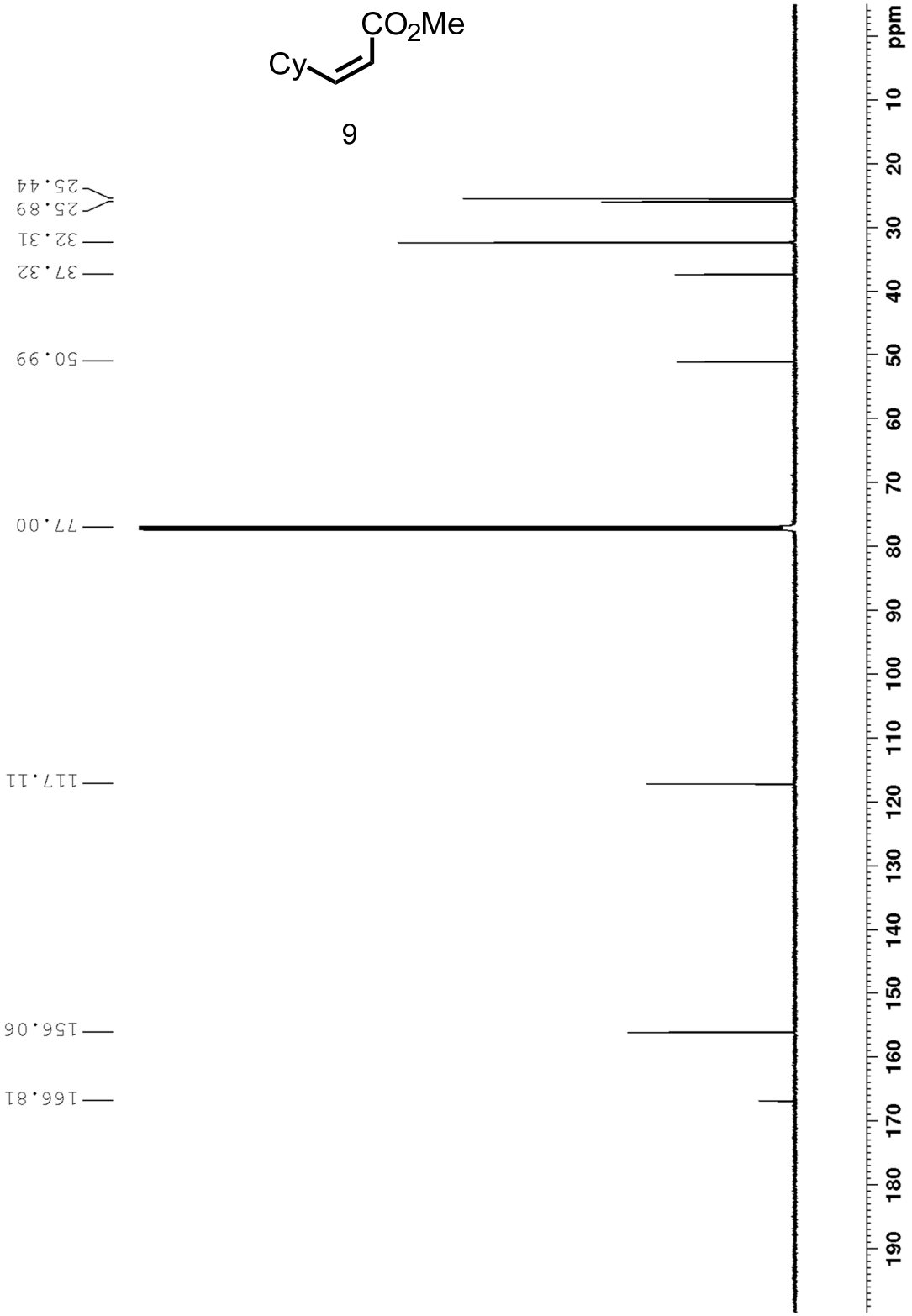


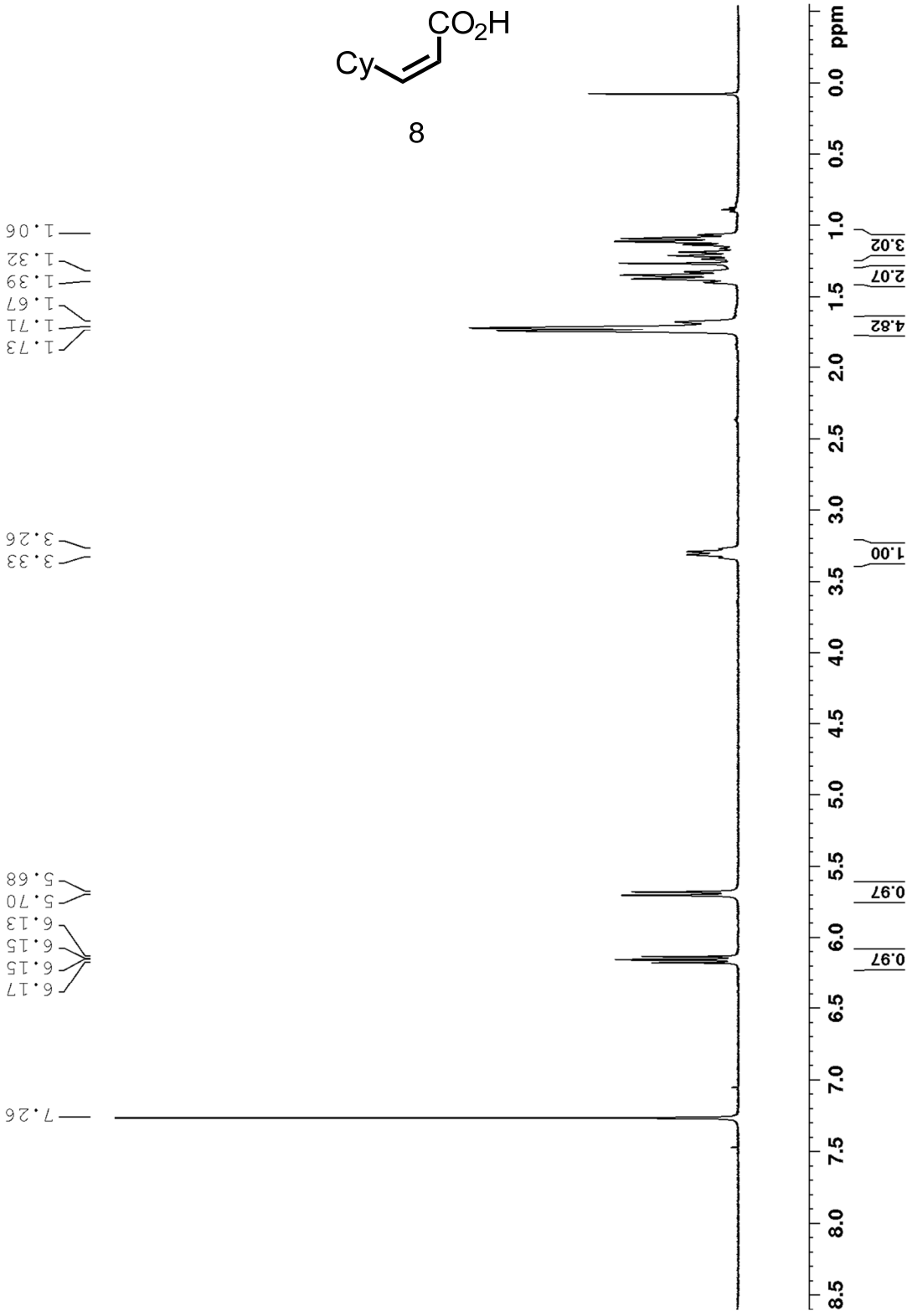
8 7

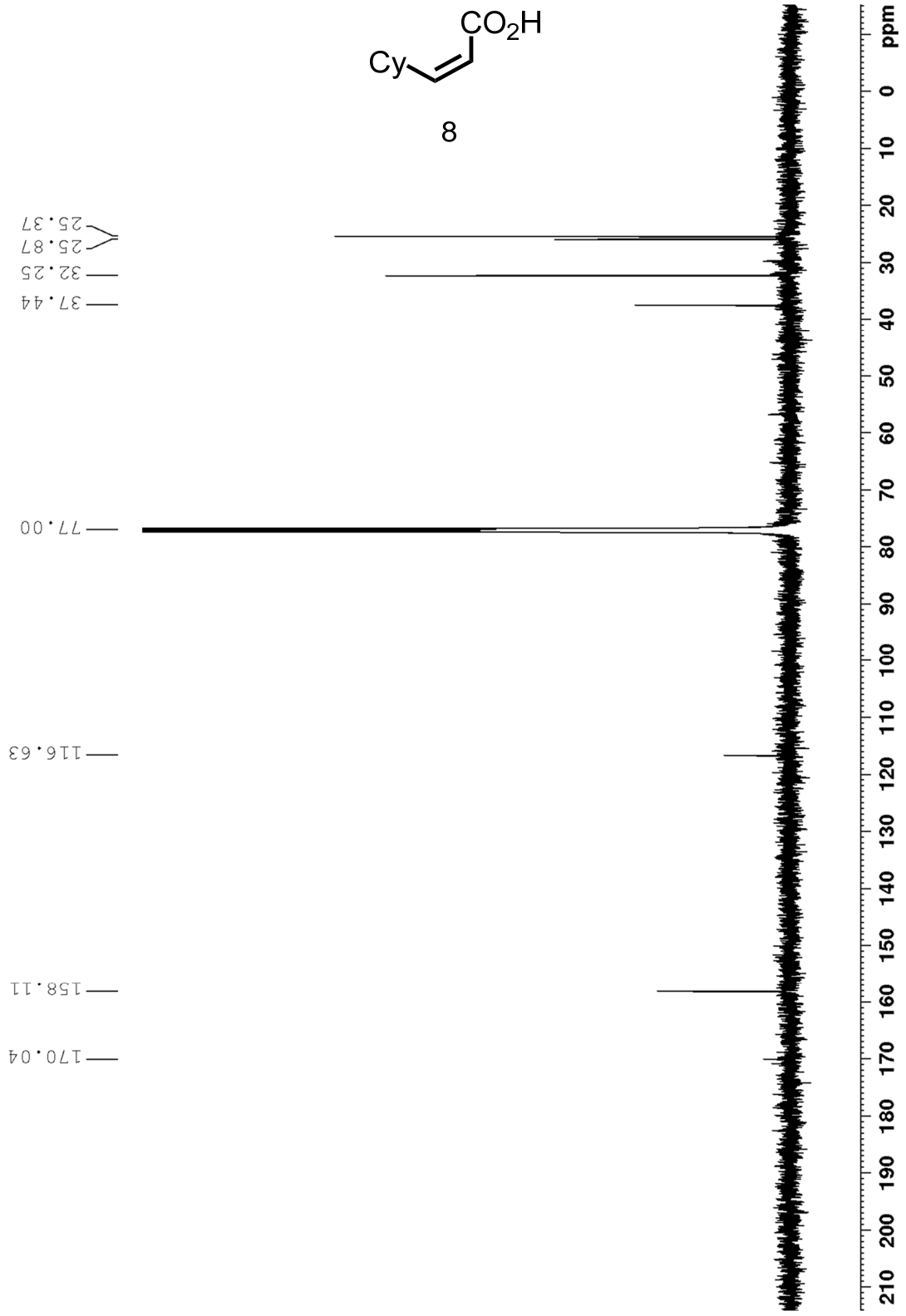
To a solution of acid **8** (20 mg, 0.13 mmol, 1.00 eq.) in dichloromethane (2 mL) at 0 °C under an atmosphere of argon were added *N*-acetylcysteamine (62 mg, 0.52 mmol, 4.00 eq.), 2,2-dimethylaminopyridine (cat.) and 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (40 mg, 0.21 mmol, 1.60 eq.) successively. The mixture was stirred reaching room temperature over night after which water was added. The aqueous layer was extracted with dichloromethane (3x), the combined extracts were dried (magnesium sulphate) and the volatiles removed *in vacuo*. The residue was purified by flash chromatography (silica gel, pentane/ethyl acetate 1:1) to give SNAC-ester **7** (30 mg, 0.11mmol, 82%) as pale yellow oil. TLC: R_f = 0.17 (pent/ethyl acetate 1:1); ^{1}H-NMR (500 MHz, $CDCl_3$): δ = 5.97 (dd, *J* = 11.3, 0.67 Hz, 1H, Cy-C*H*=), 5.87 (dd, *J* = 11.3, 9.7 Hz, 1H, Cy-CH=C*H*), 5.83 (bs, 1H, N*H*), 3.47 (q, *J* = 6.2 Hz NCH_2), 3.23-3.15 (m, 1H, Cy-C*H*), 3.08 (t, *J* = 6.2 Hz 2H, SCH_2) 1.97 (s, 3H, CH_3), 1.75-1.64 (m, 5H, Cy-CH_2), 1.37-1.29 (m, 2H, Cy-CH_2), 1.22-1.06 (m, 3H, Cy-CH_2) ppm; ^{13}C-NMR (125 MHz, $CDCl_3$): δ = 189.8 (O*C*S), 153.1 (Cy-*C*H=), 124.0 (Cy-CH=*C*H), 39.8 (N*C*H_2), 38.4 (S*C*H_2), 32.2 (2xCy-*C*H_2), 28.6 (Cy-*C*H), 25.8 (Cy-*C*H_2), 25.4 (2xCy-*C*H_2) ppm; HRMS (ESI+): *m/z* calcd. for $C_{13}H_{21}NO_2SNa$ $[M+Na]^+$ 278.1185, found 278.1187; u_{max}(film): 3283 (w), 3080 (w), 2924 (s), 2850 (w), 1672 (w), 1653 (s), 1623 (w), 1550 (m), 1446 (w), 1408 (w), 1373 (w), 1289 (w), 1262 (w), 1197 (w), 1099 (w), 1000 (s), 951 (s), 891 (w), 830 (m), 803 (w), 776 (w), 600 (w), 552 (w), 535 (w) cm^{-1}.

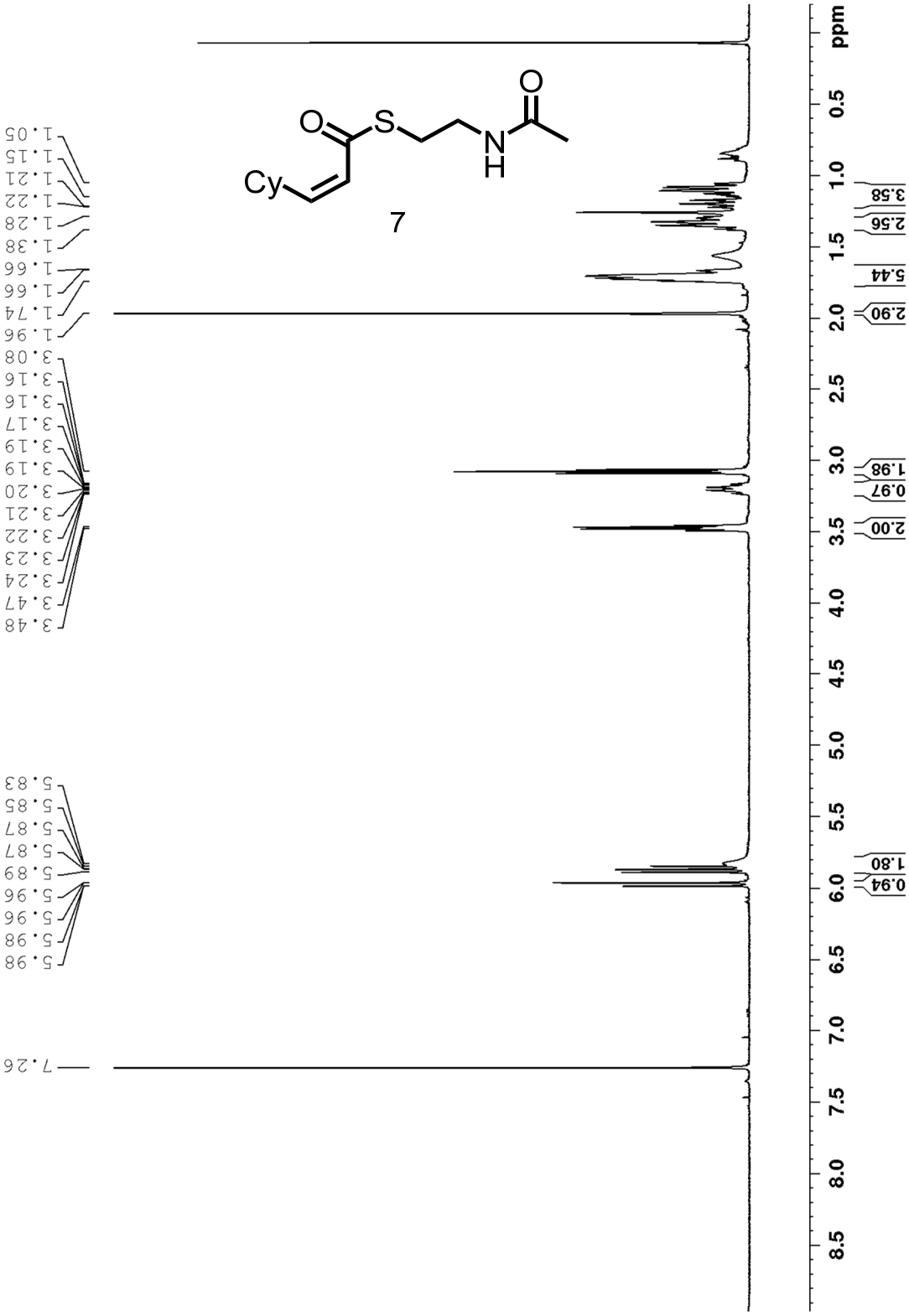
NMR-Spectra

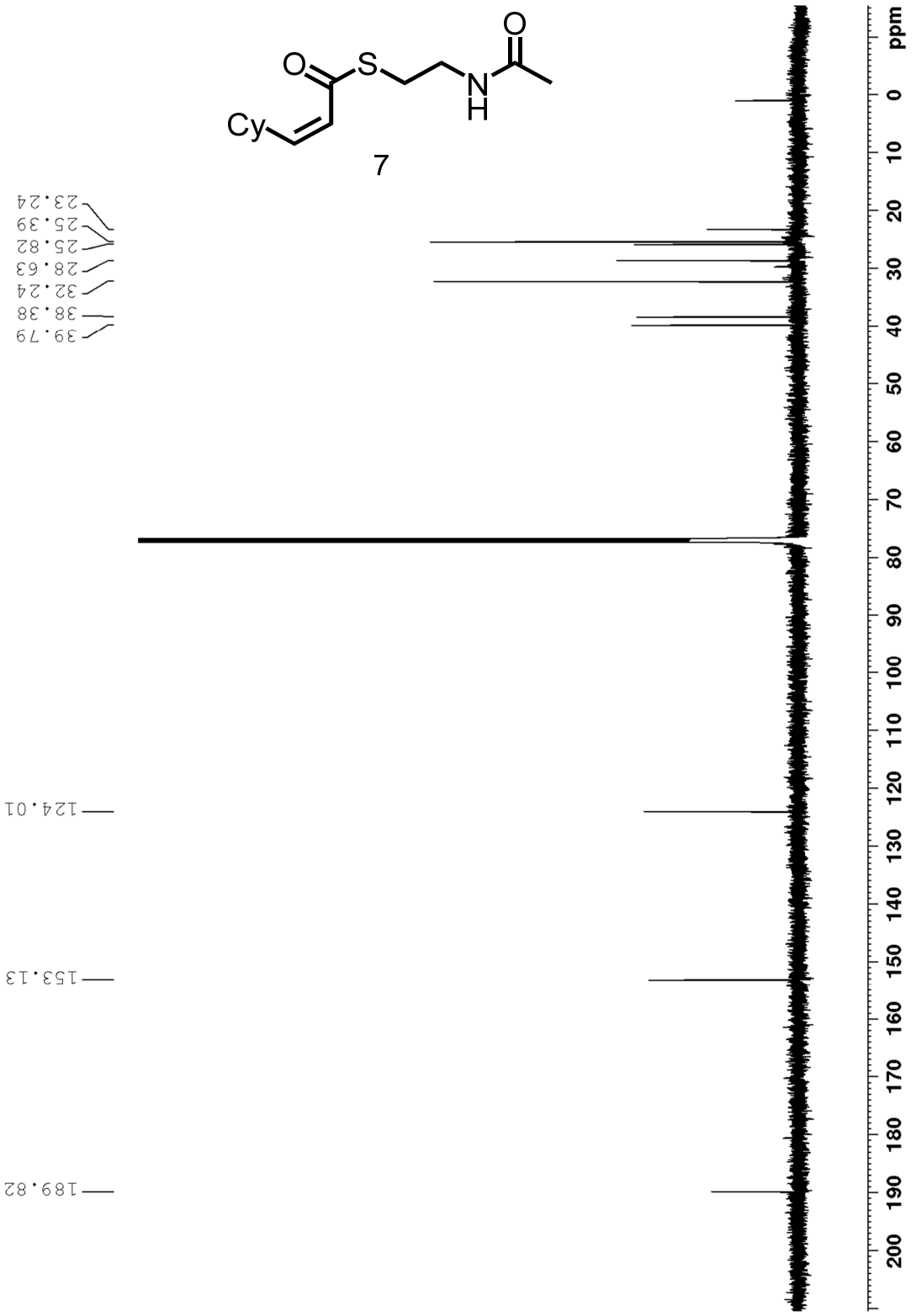

CO2Me
Cy
9
1.04
1.11
1.13
1.21
1.21
1.31
1.32
1.39
1.65
1.73
3.26
3.33
3.70
5.65
5.65
5.67
5.68
6.02
6.04
6.04
6.06
7.26
ppm
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0
2.04
1.09
2.08
5.13
0.99
3.00
0.99
1.00


CO2Me
Cy
9
25.44
25.89
32.31
37.32
50.99
77.00
117.11
156.06
166.81
ppm


CO2H
Cy
8
1.06
1.32
1.39
1.67
1.71
1.73
3.26
3.33
5.68
5.70
6.13
6.15
6.15
6.17
7.26
3.02
2.07
4.82
1.00
0.97
0.97
ppm


CO2H
Cy
8
25.37
25.87
32.25
37.44
77.00
116.63
158.11
170.04
ppm


O
S
N
H
O
Cy
7
ppm
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0
8.5
3.58
2.56
5.44
2.90
1.98
0.97
2.00
1.80
0.94
1.05
1.15
1.21
1.22
1.28
1.38
1.66
1.66
1.74
1.96
3.08
3.16
3.16
3.17
3.19
3.19
3.20
3.21
3.22
3.23
3.24
3.47
3.48
5.83
5.85
5.87
5.87
5.89
5.96
5.96
5.98
5.98
7.26


O
S
N
H
O
Cy
7
23.24
25.39
25.82
28.63
32.24
38.38
39.79
124.01
153.13
189.82
ppm
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
180
190
200

DNA sequences of the constructs used for this study. The in this study used KS-AT and ACP constructs are indicated by caption in the following DNA sequences. Marked in bold is the module linker sequence that is replaced with DEBS_TE in the cases of PnB, PnC and PnD for PnB-TE$_{DEBS}$, PnC-TE$_{DEBS}$ and PnD-TE$_{DEBS}$ constructs.

Insert	Domain	Coding DNA Sequence
PnA$_{V4}$ (PnA Accession number AFJ05078)		ATGGTAGCTGTAGCACTCGCACTGGAAACCGCTGATGCGGAACATTACGGCCCACTGTCCG ACGACGCTGAAGAACGCGTTGCAGAGCTGCGTGCACTGCGCCCTCGTGACCGCACCTCCG CTCCGGATGCTACGTCTACCTGCGCAGTACCGCTGCTGCTGTCTGGTCGCACTGACGCAGG CCTGCGTGCTCACGCAGGTCGTCTGGCAGGTCATCTGCAAGAACGTACAGACATCTCCCTG TCTGATGCAGCATTCACCCAGGCTACTACTCGCACTCTGTGGCCGCACCGTGCAATGGTGA CCGCAGCCGACCGCGACGCTGTTGTTGCTGCGCTGCGCGCTGTTGCCGAAGGTCGTGCTT CTGACGCAGCTGTACGCGCTGGTAGTGGTGGTCCGGAATCCGCTGTTTTCGTTTTCCCGGG CCAGGGTGCGCAGTGGGTAGGCATGGCACGTGAACTGGCTGCTGCCGCTCCGGTATTCCG TGACAAACTGGCTGAGTGTGCTCGTGAACTGCGCCCGTTCGTTGACTTCGAACTGGACGAC GTTCTGTCTGGTGCCCTGCCGCTGGAACGTGTTGAAGTTATCCAGCCGGCTCTGTTCGCCG TAATGGTTAGCCTGGCGGAACTGTGGCGTTCCAACGGCGTTACGCCGGCTGCTGTTGTGGG TCACTCTTTCGGTGAAATCGCTGCGGTTACTGCGGCAGGCGCTCTGACACTCGCAGACGGC GCGCGTCTGGTTGCTGCTGTATCCAAAGCTCTGGCACAACTGCAAGGTAACGGCGAGATGG TTGCAGTAGCGCTGCCGGATGCTCAGGTAACCGAACTGGTTGCAGAATGGGACCTGGACCT GGACATCGCAGTTGTAAACGGCCCGCGTTCTACCGTGGTAGCTGGCCCAACCGAAGCAGCT ACCGCTCTGGTTGAACGCCTGCGCGACCGCGATGTACGTGCAACTCTGCTGCCTATCGGTA TCGCTGGTCACAGCCGCCGTATGGAACCGGCACACGCATACCTGGTACAGGAGGCGTCTGC TGTCCGTCCGCGTGCTACTGGTATCCCAGTATACACATCTACCACTACCGACCCGCTGGACA CTGGTGACCTCGACGCCGAACACTGGTTTCATTCTCTGCGCGAACCAGCACGTTTCCAGCA GGTTATCGAAGAGCTGCTGGGTCAGGGTCACCGCGTATTCGTAGAAATGTCCCCGCATCCG GTTCTGGCTCTGTCTATCGAAGAAACCGCAGCTCACCTGGGTCGTGATGTAGTAGTTCTGGA AACGATGCGTCGTGACGATGCGGGTCACGACCGTTATCTGCGTGCTCTGGCGGAAGCACAT CTGCACGGTGTTGCGCCGGACTGGTCTACTGTTCTGCCGGATGCTCGTCGTGTGACTCTGC CGCCGTACCGTCTGGACCTGGACACAGCTGACGCTGCTGGCTCTGGCGGCACCGAACCGG GTGCCGGTCTCCGTGAACGCCTGCTGGGCGTTGCACCTGAACGTCGCATCGAAGAAGCCGT GCGTCTGGTTGTTGACGCTGTGGCAGCACACATCGAACCGTCTGCTACAGCTATTGGTGCA GACCAGGCATTCCGTAGCTTAGGCGTTGACTCCGCTGGCGCACTGCGTGTGCGCAACCGTC TGGTAGAAACTACCGGCCTGCGCCTGCCGGCGACTATCCTGTTCGACCACCCGACCCCGCG TGCTCTGGCCGAAGAGCTGGTTCGCGCTGTACTGGGCGGTGACGAACCGGCACCGGCAGC TACCGCTCCGGCTGACCCGGACGAGCCGCTGGCAATCGTAGGTATGGCGTGCCGTTTTCCA GGTGGTGTGGGTAGCCCGGACGAACTGTGGCAGCTCGTACACGAAGGCGGGGATGCCGTT GCGGCCTTCCCGGCAGATCGTGGATGGGACCTGACCGCTCTGTACGACTCTGATGCCTCCC AGCCGGGCACCTCTTACCAGCGTGAGGCTGCATTACTGGACGGTGTAGCTGAGTTCGACGC TGCATTCTTCGGTATCTCTTCTCGTGAGGCTCTCGCAATGGACCCACAGCAGCGTCTGCTGC TCGAAACTTCTTGGGAGGCTCTGGAACGTGGCGGTATTGACAGCACTACCCTGCGTGGTTC CCGCACTGGCGTTTTCACTGGTGTTATGTCTCTGCCGTACGGCCGCCCGCTCCATCAGGCG CGTCCGGACCTGGAGGGTTACGTAATGACTGGTACAACTTCTTCCGTGGTTTCTGGCCGTCT GTCTTATGTTTTAGGTCTGGAAGGTCCGGCTCTGACTCTGGACACCGCTTGCTCTAGTTCTC TGGTTGCTCTGCACTTGGCAGGTCAGTCCCTGCGTCGTGGTGAATCTGACCTGGCTCTGGT CGGCGGTGCAACTGTGATGGCTGAACCGGGCCTGTTCATCGAGTTCTCTCGTCTGCGCGCT CTGGCTCCGGACGGCCGTAGCAAGCCATTCAGCGCGGACGCTGACGGTTTCGGCATGGCA GAGGGTGTAGCTGTTCTGGTTGTTGAACGCCTGTCTGACGCACGCCGTCTGGGTCACGACG TACTGGCCGTAGTTCGCGGTTCCGCTGTGAACCAGGACGGCGCTTCTAACGGTCTGACCGC ACCGAACGGCCCGGCTCAGCAGCGTGTAATTCGCGCTGCTCTGGACTCCGCAGGTCTGCGT GCGGCTGACATCGATGTGGTTGAAGCACACGGCACAGGCACCCGCTTAGGCGACCCGATC GAAGCTCAGGCTCTGATCGCTACTTACGGTAGCGACCGTCCGGCAGAACGTCCGCTGTACG TTGGTTCTCTCAAATCTAACATCGGCCACGCACAAGCAGCAGCTGGTGTTGCTGGCGTTGTT AAGATGGTTCAGGCCCTCCGTCATGGTGTGATGCCGCGCTCTCTGTATGCGGACAACCCGA CCACCAAAGTTGAGTGGTCTTCCGGCCCGGTAGAACTGCTGGCTGAAGCACGTGAATGGGA GCGCGCAGACGGTCGCCCGCGTCGTGCAGGCGTTTCTGCCTTCGGTATCTCTGGTACTAAC

		GCACACGTGATCCTGGAAGAGTTCGCTCCGGGTACTCCGCACATCTCTGAGGGTGCTGAAT CTGCTGTTCCGTGGGCGGACGTTGCAGTTCCGCTGGTGCTGAGCGGCAAAACCCCGGATG CGGTTCGCGGTCAGGCTCTGGCACTGCGTGAACACCTGGCTTCTCACCCGGAGCTGCCATT ACCGCACGTTGCGCGCGAACTGGCTACCTGCCGTACCCGTTTCGAGCACCGTGCAGTAGTT ACTGGTGACCGTGAGGCCGTGCGTGCAGGTCTGGACGACGTTAGCCCAGTTCTGGCTGGCT CTGGCCGCGTAGCGGCAGTGTTCTCTGGTCAGGGTGCGCAGCGTGCAGGCATGGGTCGTC AGCTGGCGGACCATTTCCCGGTTTTTGACGAAGCTCTGAAGGAGGTATGTGGTCACCTCGA CCCGGAACTGGGTCGCGGCCTGCGTGAAGTTATGTGGTCTGATGACCCGGAACAGCTGGGT CGTACCGAGTTCGCTCAGCCGGCTCTCTTCGCATTCGAGGTAGCACTGGCTCGTCTGTGGC AGTCTTGGGGCGTGTCTTTCACCGCAGTTGTTGGTCACTCCGTAGGCGAAATCGCGGCCGC TGTGGTGGCAGGTGTACTGACCGTGCCGGATGCTGCGCGCCTCGTAGTTGCTCGTGGTCGC CTGATGCAGAGCCTGCCGGGCGGCGGTGCCATGTTAGCAGTCGCTGCCGGTGAGGAGGAA GTTACCGCAACTCTGGGCGACTCTGCTCTGGTTGGTATCGCGGCTGTTAACGGCCCAGAAG CAGTTGTGGTTTCTGGTGCTCGTGACGAAGTAGCTCGCGTTGGCGGCATCTGGCGTGAACG TGGTCGTCGTGTGGCTGAGCTGCGCGTTTCCCACGCTTTCCACTCTCCGCTCATGGAGCCA ATCCTGGACGAGTTTCGTGCAGTTGTAGAAAAGCTCCCGTTCCGTCCGCCAACTGTGCCGAT CATCGCTACCGCAGACACCGCGCACCCGGTTGACACCGCTGCATATTGGGTAGATCACGCG CGTAACGCTGTACGTTTTGGCGACGCAATCGGTCGCATCCCGGACGCTGACCTGCTGATCG AAGTTGGCCCGGACGCAGCTCTGGCACCGCTGATCGAAACTGGTCACACTGTACTGCCGTC TACCCGTCGCGGTCGTTCTGAAACCCACGCGGTGGTTACCGCTCTGGGTCAGGCACACGCA CACGGTGCAGACGTGGACTGGGCGGCACTGCTCCCGCCGGCTCCACGCACCGATCTGCCG ACCTACGCATTCCAGCGCCAGCGTTACTGGGACTCTGCTTCTGACGCCACTGCGCGTGTAC GTGCAGGTGCTGACCCGCGCCCGCACCCGATGCTGACCTCTCGTACTGACCTGCCGGGTG CCGGTGGTCTGCTGCTGTCTGGTCGCCTGACTCCGGGCTCTGACCCGTGGCTGTCTCATCA CGTTGTAATGGGCACCATGCTGCTCCCGGGCACCGGCTTCGTTGAGCTGGCACTGGAGGCT GCACGCTCTGCGGGTGCAGGTTCCGTTGATGAGCTGGTACTGCGCGCTCCAATGGTGTTCA CCGAAGGCCGTCCGCGTGACCTGCAAGTTTGGGTTGGTCCGGACCAGGGCGATGAACGCG AACTGCAAATTCGTACACGTGAGACTGGTGGTGACTGGACTCTGCACGCCACTGGTCTGCT GGCTGCACGTACAGCAGACACTGAAGGTTTCCGTGATGGTGACTGGGCAGGTGAGGTATGG CCACCGGCTGGTGCCCAGCAGCTGGTCGGCTCCTCCTTCTACGAAGAACTGGCTGCTCGCG GTTATGAATACGGCCCGGCATTCCACGGTACGAAGGCTCTGTGGGAGCGTGACGGTGACCT GTTTGCTGAGGTTGTACTGCCGGAAGGTCAGCCACGTGGTTTCGGTATCCATCCGGCACTG CTGGACGCTGCACTGCACGCGCTCCCGATCACCGGCTCTCTGTACGAAGCAGGTGAAGTTC GCCTGCCGTTCTCCTTCAACTCTGTGTCTCTGTTCTCCTCTGATGCACGCCGCCTGCGTGTA CGCATTCGTGCTGAGGCTGACAGCGCAGCTGTTTGGATCACCGACGATGCGGGCTCTCCGG TTCTGGCTATGGAGGGCCTGATCCTGCGCAGCATTGAACGCGCACAGCTGGAAGCAGCTGG TGCTACTGGGCGTACTGGTTGGTTCTCTGTGACCTGGCGCCAGCTGGCACGCGCAGCTACC ACCGACCGTGTTCCCGGTAACTGGCTGCTCCTGGGTGAAGTCCCGCCGGCTCTGGGTTCTC TGTTCGACGATCCGGTTACCGCAGCGTCTTGGGATCGCTCTACTGCTCCGGACGGTGTTCT GGTGGGCGCTGGTCGTGCAGAAGACCTGCTGGCTGCTCTGCACGAAGTAGCTGGCCACCG TACCGGCCCAGTATGGTGCGTAACCTCAGGTGCAGTAACTGTTGGTACTGACGACCCGGCA GCAGACGTACGTGCTGCCGGTGTTTGGGGTCTGGGTCGTGTAGCAGGCCTGGAACTGCCG GATCGTTGGGGTGGCCTGGTTGACCTGCCGGAACGCATCGACGACGCTACCCGTCGCGCA CTGGCTGGCATTCTGACTGGTGACGATGAAGGCGAAGGTGCTGACGACGCGGCAGAAGAC CAGCTGGCGCTGCGTGACGGTTGCCTGTGGGCACGCCGTCTGGTAACTGCTTCTGCCCCG CAAGCCCGCACCTGGACCCCGAAAGGCACTGTTCTGATCACTGGTGGCACGGGTGGCCTG GGTGCGCACGTTGCCCGCCGTATTGCGGAACAGGGCTCTGCTGATCGTATCCTCCTGCTGT CTCGCCAGGGTCCGGCAGCTCCGGGTGCGACCGAACTGCTGGCAGAAATCAGCGCTTTCG GCGCAACTGCTGAAGCCGTAGCTGCCGACGTGACCGACCGCGCAGCAATGGGCGGTCTGC TGGCAACTCTGGCAGCTGAAGGTGCTCCGGTTCGCACCGTTGTACATGCAGCGGGCGTTGT TCGTGACGTGCGTATTGTTGAGACTGGTGCAGAAGCTCTGGCGGCTCAGATGGCAGCAAAA GTTGAAGGCGCACTGCTGCTGGATGAACTGCTGCCGGACCTGGACGATTTCGTTCTGTTCTC CAGCATCTCTGGTGTATGGGGTGCGGCTGGTCAGGCTGGTTATGCTGCCGGTAACGCTTGC CTGGATGCGCTCGCTCGTCGCCGTCGTGAGCAGGGTCGCCCGGCTACCAGCGTTGCATGG GGTCCGTGGGCTGGCGGTGGTATGCTGACTGAACACGACGAACGTGAACTGCGTAAACGTG GTCTGACTCCGCTGCTGGTTCCGGCTGCTCTGCAAGCAATGGAACAGTCTATTATGGCAGAC TCTGCGTGGGACCCGGTAGTGGCGGACATCACCTGGTCTCGTTTCCTGCCGGCCTTCACCG TATCTCGTCCGTCTCCGCTGCTGGGCAGCTTCGAAGAAAAAACTGCACCGGGTCCGCTGGC TCCGTCTGATGGTCGTCGTACCGAAGGCGAGGCCCAGTCTCTGACTCACCGTCTCGCAGCA CTGCCGGACGCTGAACGTCTGCCGGTGCTGGCGGAAGTTGTGCGCACCCACGTTGCTGCT CTGATCGGCGAATCTGGCCCGGAACGTGTTGGCCCTGATCGTGCTCTCAAGGAAATCGGTT

		TCGATAGCCTGATGTCTGTGGAGCTGCGCAACCGCTTCGCTGGTCTGATCGGTGTGAAACT GCCGGCTACCCTGGTTTTCGACCATCCGACTCCGAACGCACTTGCTGTACACCTGCTGGGT CACCTGGACCTGGGCGCAGCTGGTACTGCCACCGCAGAGAAACCGCTGCTGGAAGTTGTTG AACAGATCGAATCTCGCGTTTTATCTCCGCTGACCGACACATCCACTCGTAAGGCACTGGCT GGTCGCTTAACTGAACTGCTGGATCGTCTGGCTGCTCTGGATGCTCGTCCGGTTACTGGCA CCGATGGTCTGGCCTCTGCAAGCGCAGCTGAACTGATGCAGTTCATCGACACTGAACTGGG TGACATCTCA
PnB (Accession number AFJ05079)	**KS-AT 1**	ATGACCACCGATGAAAGCCAGCTGGTTGAATATCTGCGTAAAGTTACCACCGATCTGCAAAA AACCCGTAGCCGTCTGCGTGATGCAGAAACCAAACATCATGAACCGATTGCCATTGTTGGTA TTGGTTGTCGTTATCCGGGTGGTGTTCATGGTGCAGATGATCTGTGGCGTCTGGTTGTTGAA GGTCGTGATGCAATTAGCGGTTTTCCGCAGGATCGTGGTTGGGATCTGGCAGGTCTGTATG ATCCTGATCCGGATGCAGTTGGCACCAGCTATACCCGTGAAGGTGGTTTTCTGACCGGCATT GACCTGTTTGATGCAGATTTTTTTGGTCTGAGTCCGCGTGAAGCACTGGCAATGGACCCGCA GCAGCGTCTGCTGCTGGAAACCAGCTGGGAAGCACTGGAAAGCGCAGGTATTCGTCCGGCA ACCCTGCGTGGTAGCCGTACAGGTGTTTTTACAGGTGTTATGCATCATGATTATGGTAGCCG TTTTAGCAAAGCACCGGATGGTTTTGAAGGTTATGTTCACATGGGTAGCGCAGGTAGCATGG CAGTTGGTCGTGTTGCATATCATTTTGGTCTGGAAGGTCCGGCAGTTACCGTTGATACCGCA TGTAGCAGCAGCCTGGTTAGCCTGCATCTGGCATGTCAGAGCCTGCGTACAGGTGAATGTG ATCTGGCACTGGCAGGCGGTGTTGCAATTATGGCAGGTCCGACCTTTTTTGTTGAATATAGC CGTCAGCGTGTTCTGAGCACCGATGGTCGTTGTCATAGCTTTGCAGATGATGGTGATGGTAC AGGTTGGAGCGAAGGTGTTGGTGTTCTGGCAGTTGAACGTCTGAGTGATGCACGTCGTCTG GGTCATGATGTTCTGGCCGTTGTTCGTGGTAGCGCAGTTAATCAGGATGGTGCAAGCAATGG TATGACCGCACCGAGTGGTCCGGCACAGCAGCGTGTTATTGATAGCGCACTGCGTAATGCA CGTCTGAGCGCAGATGAAATTGATGTTGTTGAAGCACATGGCACCGGCACCCGTCTGGGTG ATCCGATTGAAGCACAGGCCCTGATGGCAACCTATGGTCGTGGTCGTAGCGCACAGCGTCC GCTGTGGCTGGGTAGCCTGAAAAGCAATATTGGTCATACCCAGGCAGCAGCCGGTGTTGGT GGTGTGATTAAAATGGTTCAGGCACTGCGTCATGGTGTTCTGCCTCGTAGCCTGTATGCAGA ACGTCCGAGCACCGAAGTTGATTGGAGCGCAGGCACCGTTGCACTGCTGGCACGTGAACGT GCATGGGAACCTGTTGAAGATCGTCCGCGTCGTGCCGGTGTTAGCAGCTTTGGTATGAGCG GCACCAATGCACATGTTATTGTTGAAGAATTTGTTCCGGAACCGGCAGGCGAAAGCGAAGCA GAACCGAGCGCACCGTGGGCAGAAAATAGCGTTCCGCTGGTTCTGAGCGGTCGTACACCG GAAGCAGTTCGTGCACAGGCAGCAGCACTGCGTGATCATCTGGCAGCACGTCCGGAACTGA GCCTGGCACGTGCAGCACGTGAAAGCGCAGTTCATCGTACCGCATTTGAACATCGTGCAAC CGTTACAGGTGATCGTGATGCAGTTCTGGCAGGCCTGGGTCATGTTACTCCGGTTGCAGCC GGTGGTGGTCGCGTTGCAGCAGTTTTTAGCGGTCAGGGTGCACAGCATGTTGGTATGGGTC GTCAGCTGGCAGCAGATTTTCCGGTTTTTGCCGGTGTTCTGGATGAAGTTTGTGCAGTTGTT GATCCGCTGCTGGGTCGTAGCCTGCGTGAAGTTATGTGGTCAGGTAGCGCAGAACAGCTGG AACGTACCGAATTTGCACAGCCTGCACTGTTTGCATTTGAAGTTGCACTGGCACGTCTGTGG CAGAGCTGGGGTGTTGAATTTAGCGTTCTGGCAGGTCATAGCGTTGGTGAAATTGCAGCAG CAGTTGTTGCCGGTGCACTGAGCCTGGATGATGCAGCCCGTCTGGCAGTTGTGCGTGGTCG TCTGATGCAGCGTCTGCCGGAAGGTGGTGCAATGCTGGCAGTTGCAGCGGGTGAAGAAGAA GTTGCAGCAACCCTGGGCGATCCGGCACTGGTTGGTATTGCAGCAGTTAATGGTCCGGAAG CCGTTGTTGTTAGCGGTGCACGTGCCGAAGTTGAACGTGTTGCCGGTGTTTGGCGTGAACG TGGTCGTCGTACCAGTCGTCTGCGTGTTAGCCATGCATTTCATAGTCCGCTGATGGAACCTG TGCTGGATGAATTTCGTGCCGTTGCAGATGAACTGGATCGTCGTGAACCGACCCTGCCGGTT ATGGCAAGCGCAGAAACCACCCATCCGTTTGCAACCGCAGCATATTGGGTTGATCATGCACG TCGTGCAGTTCGTTTTCATGATGCAGTTACCCGTCTGGCCGAAGCAGATGTTGTTGTTGAAAT TGGTCCGGATGCGGTTCTGGCACCGCTGATTGATACAGGTCGTACCGTTCTGAGCAGCTGT CGTCGTGAACAGAGCGAAACCCTGACCCTGCTGACCGCACTGGGTGAAGCACATGCACATG GTGTTGAAGTGGATTGGACCACCCTGTTTCCGGCAGCACCGCGTGCCGATCTGCCTGCATA TCCGTTTCAGCATCGTCGTTATTGGCTGGCACCG
	DH-KR 1	CCTGCACTGACAGGTGCCGGTGCAGATGCCCTGGACCATCCGCTGCTGAGCAGCCTGGTG GAACTGCCTGGTCAGGGTGGTGTTGTTCTGACAGGTCGTATTAGTCCGGATCGTGATCCGT GGCTGGCAGATCATGCAGTTAGCGGTGCCGTTCTGTTTCCGGGTACAGGTTTTCTGGAACTG GCACTGCGTGCCGCACGTCAGGCAGGTTGGCGTCAGGTTGCCGATCTGGTTGTTCAGGCAC CGCTGGTGCTGCCAGCAGCGGGTGCAGATATTCAGGTTTGGGTTGAACCGCCTGGTGAAGC AGAACGTGCACTGGTGGTTCGTGGTCGTTCAGGTGATGGTGATTGGGTTGAATATGCAACAG GTCGTCTGGTTGAAGAACCGCCTGAAACCAGTCCGCAGGGTAGCCCTCCGACCGATGCAGA ATGGGCAACAGGTCAGTGGCCTCCGCAGGGTGCAGAAGAAATGCCGGTTGAAGAACTGTAT GATGCACTGGCAGCCCGTGGTTATGGTTATGGTCCGGTTTTTCGTGGTCTGCGTGCACTGTG

		GCGTAGCGGTGATGAACTGTATGCCGAAGTTGTTCTGCCGGAACGTGCACAGGATGGTCGT TATGATTGTCATCCGGCACTGCTGGATGCAGCACTGCATCCGCTGGCAGCAGCAGCAGATA GTGGTCAGCAGGTTCGTCTGCCGTTTGCATTTGGTCGTGCCACCGTTCATAGTCCGGGTGC AAGCGAACTGCGTGTTCGTCTGCGTACGGGTGCAGAAACCATGCGTCTGGATGCAGTTAGC CCGACAGGTGAACCTGTTCTGACCATTGGTGAACTGGTTCTGCGTGCAGCAGATACCGCAC GTCTGACAGCCCGTGCCGAAGGTCATGTTCATCGTCTGGGTTATGAAGTTACCTGGCAGCTG CTGCCTGATCCGCCTCGTGCAGATCGTCTGCCTGGCACCTGGATTGTTCTGGCACCTCCGG CAACCGATGTTAGCTGGACCCGTGATCTGGCAGAACATACCCTGGCAGTTGATCTGGATACC ACACCGGATCGTCCGGGTCTGGTTGCACGTCTGGCAGATGTTACCCATGTTCCGACCGCAA CCGCACCGGATGGCGTTCTGCTGTTTGCAGCACATCCTGATCTGCTGGTTACCGCACTGCAA GCACTGACCGATGCCGGTGTTGATGCTCCGGTTTGGTGTCTGACCCATGAAAGTGATGATGA GCTGGATACCGCAGCAGTTTGGGGTGCAGGTCGTGCAGCAGCCCTGGAACTGCCGGATCG TTGGGGTGGTCTGGTTGATCTGCCTCCGGCAGGCGTTCGTCCGGATCTGGGTTGTCTGGCT GGTCTGCTGAGTACCGATACAGGTGAAGATCATCTGCGTCTGCGTCCGGATGGTGTTAGCG TTCGTCGTCTGATTAAACCTGATCGTCCGACCCAGCCTCCGGCAGATTGGACCCCGAGCGG CACCGTTCTGATTACAGGTGGTACAGGTGCACTGGGTGGTCATGTTGCACGTTGGGTTGCA GGTCTGGGTGGTTGTAGCCTGCTGCTGGTTAGCCGTCGTGGTCCTGAAGCACCGGGTGCAC AAGAACTGCTGGCAGAACTGGCAGCAACCGGCACTCCGGTTCGTATTGTTGCAGCAGATGT TGCAGATCGTAGCGCAATGGCAGGTCTGGTTCTGGAAGCAGCAGATGCAGGCACCCCTGTT CGTGCAGTTTTTCATGCAGCGGGTGTTGCCGATGAAACACCTCTGCTGGAAACAACCCTGG GTCGTTTTCATGCCGTTCTGGAAGGTAAAGCCGGTGGTGCACGTGTGCTGGACGAAGTTCT GGGTGAAACCGATCTGGATGCATTTGTTGTTTTTAGCAGCGTTAGCGGTGTTTGGGGTGGTG CAGGTCAGAGCGCCTATGGTGCAGGTAATGCAGTGCTGGATGCCCTGGCAAGCCGTCGTCG TGCACGTGGTCTGGCAGGCACCGCACTGGCATGGGGTCCGTGGTCAGGTGGTGGTATGGT TGATGATGATCTGGAACGTCGTCTGCGTCGTCAGGGTCTGAATCCGCTGCCGGTTGATGAT GCCATTGCAGCACTGAGCACCGCAGTTCCGCTGGGTGAAAATCGTGTTCTGGCGGATGTTG CATGGTCACGTTTTCTGCCGCTGTTTGCCGCAACCCGTCCGGCACCGCTGTTTACAGCACTG CCTCCGACACCGCGTCATACCGCAAGCGCAGCAACAGCAGCACCGGGTCCGACCGCAGAA
	ACP 1	CGTCTGGCAGCACTGCCTGCCGCAGATCGTGCAAAAGCACTGAGCGATCTGGTTCGTGCAG AAATTGCAAGCGCAGTGGGTCATCATACCATGAGCGCAGTTGATGCCGAACGTCCGCTGGG CGAATTTGGTTTTGATAGCCTGATGAGCGTTCAGCTGCGTAATCGTCTGAGTGCAGCAACAG GTGTGCGTCTGCCTGCAACCCTGGTTTTTGATCATCCGACCGCAGGCGCACTGGCACGCCA TCTGGAAAGCGAACTGGCAACCGAACCGAGCACCCGTCCGCGTCAGCCTGCC
	KS-AT 2	ACCGCAGAAACCGCAGTTGCGGATGAACCGCTGGCGATTGTGGGTATGGCCTGTCGCTATC CTGGTGGCGTTGCAAGCCCTGAGGACCTGTGGCGTCTGGTGGCACAGGGTCGTGATGCCG TTACCGCCTTTCCGACCGATCGCGGATGGGACCTGGCTGGCCTGTATCATCCGGATCGTAG CCGTGCAGGCACCACCTATGCAAGCGAAGGTGGCTTTCTGGATGATCCGGCAGGTTTTGAT GCCGATTTCTTCGGTATTTCACCGCGTGAGGCCCTGGCTATGGACCCTCAACAACGCCTGCT GCTGGAAGCAAGTTGGGCAGCAGTTGAACATGCAGGTATTGCACCGGATACCCTGCGTGGC TCACGTACGGGTGTGTTTGTTGGCACCATGTATAATGATTATTTCAGCCGTCTGAGCAGCAC ACCGGAAAGCCTGGAAGGTATTATTGGTATTGCAAATAGCAATAGCGTTATGAGCGGTCGCA TTAGCTATCTGCTGGGTCTGCAAGGTCCTGCCGTGACCCTGGATACCGCCTGTAGCTCAAG CCTGGTGGCACTGCACCTGGCAGGTCAGTCACTGCGTCGTGGCGAATGCGACCTGGCCCT GGCTGGGGGTGCCACCGTTATGGCAAGTCCGCATATTTTTGTGGAATTTAGTCGCCAGGGT GGTCTGGCCGTTGATGGTCGTTGCAAAAGTTTTAGCGCAGATGCAGATGGCACCGGCTGGT CAGAAGGTGTGGGTCTGCTGGTGCTGGAACGTCTGTCAGATGCCCGTCGCCTGGGCCACG AAGTGCTGGCGGTTGTTCATGGTTCAGCCGTGAACCAGGATGGTGCCTCAAATGGCCTGAC CGCACCGTCAGGTCCTGCCCAGCAGCGCGTTGTGGAAGCAGCACTGGCACAGGCAGGTAT TGCCGCAACCGAAGTGGATGCAGTTGAAGCGCATGGTACAGGTACACGCCTGGGCGACCC GATCGAAGCCCAAGCACTGATTGCCACGTATGGTCGTGAACGTGAAGCCGAACGTCCTCTG TATCTGGGTAGTCTGAAATCAAATATTGGCCACGCGCAGGCAGCCGCAGGCGTGGGTGGCG TTATCAAAATGGTGCAAGCCCTGCGCCATGAACTGCTGCCTCGTACCCTGCACGCAGATACC CCGAGTCCGGAAGTAGATTGGAGCAGCGAAGCCGTTCGTCTGCTGACCGAAGAACGTCCGT GGCCTCGTGGTGATCGTCCTCGTCTGGCAGGCGTTAGCTCATTTGGTATTAGCGGTACAAAC GCCCATGTTGTTCTGGCTGAAGGTGATCCGCTGGATGGTGAACCTGCCGAAGGTGCTCCGG TTAATGGTGAAAGCGTTGAAAGCGGCACCCCGATGGATACCAGCGATAGCGCAGCAGCCCG TGCACCGCTGCCGTATGTTCTGACCGCACGTAGCGCAGCAGGTCTGCCTGCACAGGCACGT GCACTGCATAGCCATCTGATTGGTCGTCCTGGTCTGGAACCGGCAGAAGTTGCACGTAGCC TGGTTACCACCCGTAGCCTGCATGATCATCGTGCGGTTGTTGTTGCAGAAGATCGTCAAGAG CTGCTGGACGCACTGGCAGCGCTGGCAGATCCGGCAACCACCACACCGCGTCGCACCGTT AGTACCGGCACCGCAGGTCGTGGCCGTGTTGCGGGTCTGTTTTCAGGTCAGGGTGCCCAAC

		GTCCGGGTATGGGTCGCGAACTGGCAGGTCGTTTTCCGGTGTTTGCAGGCGTGCTGGACGA GGTTTGTGCCGTGGTGGACCCTCTGCTGGGACGTTCACTGCGCGAGGTGATGTGGTCAGCA CCGGGTGAAGTTCTGGAACGCACAGAGTTTGCCCAGCCAGCCCTGTTTGCCTTCGAGGTGG CTCTGGCTCGCCTGTGGCAGTCATGGGGAGTGGAATTTTCAGTGCTGGCAGGCCATTCAGT GGGTGAAATCGCAGCCGCATGTGTTAGCGGTGTGCTGAGCCTGCCGGATGCCGCACGTCT GGTTGTGGCACGTGGACGCCTGATGCAGGCACTGCCTGAAGGCGGAGCAATGGTTGCGATT GCAGCAGGCGAAGATGAAGTTAGCGCCAGCCTGGCCGATGTTCCGGATGTTGCCATTGCCG CTGTGAATGGTCCAGAAGCAGTGGTAGTTTCAGGCACCGAAGCCGATGTTCTGCGTAGCGC AGATCATTGGCGTGAACAGGGACGTCGCACCTCACGTCTGCGTGTGTCACATGCCTTTCATT CACCTCTGATGGAACCAATGCTGGACGATTTTGCAGCAGTTCTGACCGAACTGACCTTTCAT GAACCGGCACTGCCGATTAGCCCGAGCGCAGATAGCAGCCGTAGCTTTGCCAGCGCAGAAT ATTGGCTGGACCATGCCCGTCATGCCGTGCGTTTTGCAGATGCACTGAGTGGTCTGGATGG TGCGGATGTTCTGGTTGAACTGGGTCCTGATGCAGCCCTGGCACCTCTGGCAGGTACAGAT AAACCTGTTCTGGTTTGTGCACGTCGTAATCAGCCGGAAGTTCGTACCCTGGTTACAGCACT GGGTGCCGGTCATGCGCATGGTGTGGCGGTTGATTGGACAGCAGTGCTGGGTGAAGGTCG TCAGGTGCGTCTGCCGACCTATAGCTTTCAGCATGAACGCTATTGGCTGGATGAAGATACCG CAGCCGGTCGTGGTACTGGTGCA
	KR 2	GCACATGATGCAGCGCAGGCACGTTTTTGGCAGGTTGTTGAAGATCAGGATCTGGATGGTCT GACCCGTACCCTGGGTCTGGATGCCGAAACCAGCCTGCGTGCAGCCCTGCCGGTTCTGCAT GATTGGCATCGTACCAGCAGCACCCTGGCGCAGGCAGCGGGTTGGCGCTATCGTGTTGCAT GGGATCGTGTTCCGACCGATGATAGCGCAGTTGCACTGGATGGCACCTGGTGGATTGTTGT TCCGAGTGATGCAGCAGATGCCACCACCGCAGATGCAGTTCGTCGTGCCCTGGCAGCGGCA GGCGCAAATCCGCGTATTCTGACGATTGATCCGCATCGTACCGATCGTGCAGCACTGGCCA AAGAGCTGGCAGCCGCAGCCGATGGTACAGTTGCAGGTACAGTTAGCCTGCTGGCCGAAAG CGGTGGTGAAGATACAGGTCATCGTGGTGTGGCAGCCGGTGCCCTGGCCACCCTGGTTCTG CTGCAAGCGCTGCATGATGCAGATATTACCACCCGTCTGTGGACCCTGACCCGTGGTGCAG TTCGTACAGGTCCGGCAGATACAGCTCCGGGTCCGTGGCACGCACAGGTTTGGGGTCTGG GTCGTGTTGCCGCACTGGAACATCCTGCACTGTGGGGTGGCCTGATCGATCTGCCAGCAGA AGGTGAACCGGCAGGTCTGGCAGCCGTTCTGAGTGGTACAGCCGGTGAAGATCAGGTTGCA CTGCGTGAAGATGGTGTTCGTGCCCGTCGTCTGACACCGGCAGATACCCATGATGCACCGG GTGCCGATGCACAGGCACCGGGTGCTGATCGTTGGACCGATGGTGCAGTTCTGATCACGGG TGGCACGGGTGCCCTGGGTGCACATACCGCACGTATGCTGGCCAGCCGTGGTGCAAGCCG TCTGGTGCTGGTTTCACGTCGTGGTGCAGCCGCACCGGGTGCGGAAGCACTGCGTGCGGA ACTGGAAACCCTGGGTGCAACCGTTGATCTGACCGCATGTGATGTTACCGATCGTGATGCG GTTGGTCGTCTGGCAGAAGCCCTGGCAGCCGAAGGTACACCTGTTCGTGCCGTGGTTCATG CCGCAGGCGTTGCAGCGGAACGTCCACTGACAGAACTGGTTGGTGATGATTTTTCAGCAGTT AGTGATGCAAAAGTTACCGCAGCAGAAATTCTGGATGATGTGCTGGGTGATGACCTGGCAG CCTTTGTTGTGTATAGCAGCATTGCAGGCACCTGGGGTTCAGCCGGTAATGGTCCGTATGCA GCCGCAAATGCACATCTGGATGCACTGGTTGAACGTCGCCGTGCCCGTGGTGCCGTTGGCA CCGCTATTGCCTGGGGTCCTTGGAGCGGTGGCGGTATGGCAGATGATCGTTTTCAAGAAGA AATGCTGCGTCGCGGTGTTAGCGCACTGTCACCGCAGGGTGCCACCGCAGCACTGGCGCA AGCCCTGGAACATGATGATACCACCGTTACCGTTGTTGATGTTGATTGGGATCGTTTTGCAC GTGTTTTTGCATGTAATCGTCCGAGTCCGCTGCTGCGTCATCTGACCGAACCTGCCGCAGCT GCCGGTGAAGCACCGGCACGTACCGAACTGGCAGAG
	ACP 2	CGGCTGGCAGCCCTGGATGCAGGTACACGTCGTGGCGCAGTTCTGGATCTGGTGCGTGCG GAAGTTGCGGGTGTTCTGGGTCATGCAACCAGCCAGGCAGTTGCAGTTGATCGTGCATTTAC CGATATGGGTTTTGATTCACTGATGGCCGTTGAACTGCGTGCACGTCTGGGTGCCGTTACAG GTCTGGCACTGCCGACCACCCTGGTGTTCGATCGTCCGAGCCCTGCCGAAGTTGCAGGTTT TCTGTGTGATCGTTTTGAA**CCAGATCGTGATGCCCTGACACATGAAATTCTG**
	linker	**GAAAAACTGGATTGGCTGGAAAAAACGCTGCTGGATGCTGCGCATAGCCAGGGTGCCCGT GCACGTTTTGGTAGCCGTCTGGATGCGCTGCTGGCACGCCTGGATCGTACCACCGTTCCG CCTGAAAGTCGTCTGAGCGATCCGGATACCGATACCGATGCCATTGAAAGCGCAACCGCA GAGGAACTGCTGGCCTTTGTTGAAAAACACTTTGATTCA**
PnC (Accession number AFJ05066)	**KS-AT**	ATGGTTATGACAGGCGAAGGTCGTAGCGATCAGCTGGTGACCGCACTGCGCAAGGTAACTC GTGACCTGCGTGACACCCAACGTCGCCTGCAAGTTGAAGAGGACCGTAACGGTGAGCCAAT CGCTGTCGTTGGTATGAGCTGCCGTTTCCCGGGTCACGTCGACTCTCCGGAAGCACTGTGG GAACTGGTTTCTTCCGGCCGTGGTGCTTCTGGTTCTTTCCCGGTTGATCGTGGTTGGGACCT GGACGCTCTGTTCTCCGAATCTGCTGACGGTAAAGTTCCGTCCTACGTTCGTGAAGGTGGCT TCCTGTATGATGCTGGTTGGTTCGACGCTTCTTTCTTCGGCATTTCTCCGCGTGAGGCGGTA ACTATCGACCCGCAGCAACGTCTGCTGCTGGAAGTTGCTTGGGAATCCCTGGAACGTGCAC

		GCATCAGCCCGCAGGCACTGCGTGGTTCTGACGTAGGTGTTTTCGCTGGTGCAATGAACCA GGACTACGCAGTTCGCCTGCATGAAAGCATCGAAGATTTCGAAGGTTTCCTGACCACCGGCA ACACCGGCTCTGTAGTTTCCGGTCGTCTGAGCTATACGCTGGGTCTGGTAGGCCCGGCTGT TACCGTAGACACTGCTTGTTCTTCCTCTCTGGTAACCGTGCACATGGCAGCACGCTCTCTGC GCTCTCGTGAATGCTCTCTGGCCCTCGCAGGCGGCGTAACCGTAATGTCTATGCCGACCAC CTTTACTGAGTTCTGCCGTCAGCGCAACCTGGCGCCGGACGGTCGCGCTAAGTCCTTCGCA GCAGCAGCTGATGGTACTGCTTGGGCTGAAGGCGCAGGTATGCTGGTGCTGGAACGTTTAA GCGACGCTGAGCGTAACGGTCATCCGGTGCTGGCTCTGATCCGTGGGTCTGCTGTAAACCA GGACGGTGCTTCTAACGGCCTGTCTGCTCCGAACGGCCCGTCTCAAGAACGTGTGATCTGG CAGGCCCTGCGCGATGCTCGTCTGACTGCTGACGAGATCGATGCAGTTGAAGCTCACGCAA CCGGGACTCGTCTGGGTGACCCGATCGAAGCTCAAGCACTGCTGGCAACCTACGGTCAGG GCCGTTCTGCTGACCGTCCGCTGTGGCTGGGTTCTCTGAAATCCAACATCGGTCACGCACA GGCAGCTGCTGGTGTGGGCGGTGTTATCAAGATGGTAATGGCTTTACGCCACGGCGTTCTG CCGCAGACCCTGCACGTGGACGCGCCGACCCCGGAAGTTGATTGGTCTGCTGGTGAGGTT CGCCTGCTGACGGAGCGTCGTGAATGGCCGCGTGCAGGCCGCCCGCGTCGTGCTGGCGTA TCCTCTTTTGGTGTGTCCGGTACTAACGCTCACCTGATCCTGGAAGAGGCGCCGGCAGCTC CGCCTCGTGACGGTGCAGCTGACCGTGCTGACGATGGTGGTGGCGCTACCCTGCCGTGGG TTCTCTCTGCACGTACCGCAGCTGCTGTGCGTGAGCAGGCTCGCCGTCTTCATGGTCACCT GACTGACCATCCGGAGCTTGTTCCGGCACAGGTTGCTCGTTCTCTGGTAACCACTCGTTCCA CCTTCGAACAGCGCGCAGTTGTTCTGGGTACTGACCGCGCTGAGCTGCTGGACGGTCTGGA TTCTATCGTTCAGGGTGCCCCGGACGCACGTACTGTTACCGGCTCCGCTGTTGGTGGTCGT GAGGTAGTATTCGTGTTCCCAGGGCAGGGTGGTCAGTGGGCTGGCATGGCACTGGAACTGA TGGAAGAATTTCCGGTTTTCGCTGAAACCCTGCACTCTTGCGCTGACGCCCTGGCTGACTAC GTTGACTGGTCTCTGCTGGACGTACTGCGCGAGGCAGAGGGTGCTCCGGGTCTGGACCGT GTTGACGTCATCCAGCCGGTTCTGTTCTCTGTAACCGTTGCACTGGCAGACCTGTGGCGTAG CCTGGGCATCGAGCCGTCTGCTGTCGTTGGTTCTAGTCTGGGCGAGATCGCAGCTGCATAC ACTGCTGGTACTCTGAGTCTGGCAGACGCGACCAAAGTGGCCGTTCTGCGCAGCCGTGCAC TGCTGGAACTGTCTGGTCGTTCTGGCATGGTTTCCGTGCCGCTGGGTCGTGCTCAGGTAGA AGAACTGATCGCAGCTTGGCCGGATCGCCTGTTCGTGAGCGGTGTTAACTCTCCATCTATCT CCGTAGTTTCTGGCGACAACGAGGCAGTGGACGAACTGCTGGCGGTTTGTGCTGAGCGTGG TGTTCGTGCACGTCGTGTTGCTACCGACTGTGCTTCTCACTACCCGGCTGTAGAAGCTCTGG AAGAGCGTCTGCTGACGGACCTGGCAGACCTGGCTCCGACTCCGGGTCGTATCGCTTTCCT GTCCACCGTTTCTGGTGAAAGCGCAGAAACTGAACCAGTACCGGATGCTGCTTACTGGTATC GTAACACTCGCCGCACCGTTGAGTTCGAGCCGGTTATTCAGCGTCTGGCAGCGACTGGTTC CAAAGTTTACATCGAAATCTCTCCACATCCGGTGCTGCAAGTTGCTCTGAGCGAGATCGTGG AAGGTGAATCTCGTGAAGCTGCGGTTCTGAGCACTCTGCGTCGTAACACTTCTGACCGCCG CGCTTTCCTGACCAGCCTGGCGAAAGCATACGTTTCTGGTGTAACTGTGGACTGGGCCGCA CTGCCGTCCCTGGCGGGCGCAGCACAGGTGGACCTGCCGACTTATGCATTCCAACGTGAGC GCTACTGGCCGCGCCCGGCAGCTGCAAGTAACGGCGGTCACGGT
	KR	GCTGGTACAGGTGCTCGTGCTGGTGTTGGCCACGGTACTGTTGACGCCCACTTCTGGGAAG CTGTAGAGAACGGCGACCTGGGTAGCCTGGGTCCAGACGTACGCTTTGACGACGAAACCCC GCTGAAGGAGGCACTGCCGGAACTGGCTTCTTGGCACCGTCACGGTCTTGAACAGGCACGC GTTGACGGTTGGCGTTACGTTGAACGTTGGCGTCCACTGGACGTTCCGGCTACTGGTCCAC ACGGCAAATGGCTGCTGGTGCACCCGGGTGGTACTGAATCCGATGCGACGACGGAATGGG TACGCGAATCTCTGCTCGAGGCTGGTTGTTCTGTAGTGGACCTGCTGGTTGACACTACTGAC ACTGACCCGGCAGCTGTGACGGAGCGTCTGAACCAGGCTTGCGCTGGTGGTGGCCCGGAG CCGGTTGGTGTGGTTAGTCTGCTGGCGTTCGACGGCCGCCACGACACTGCTCGCCCGTCC GTACCGCGTGGTACAGCAGCTACTCTGGCTCTGGTACGTGCACTGGGTGCAGCGGGTATCA CTGCTCCGCTGTGGTGCCTGACCCAGGGTGCAGTTACAACTGGCACGGGTGACCGTCTGGA CGCTGTAGAACAGGCACAAATCTGGGGTCTGGGCCGTGTGGTTGCACTGGAACACCCGGAT CGTTGGGGCGGTCTGGTTGACCTGCCGTCTACACTGGACCCGCACGTACGTACTCAGCTGT GTGCAGCACTGGGCGGTGCTCATACCGACGAGGACCAGCTGGCCCTGCGTCCGGTGGGCA TCCTGGGTCGCCGTCTGGTTCCGGATGACGGTGACCGCCAGGCAGCTCGTACTGCACCGT GGCGTCCAGCTGGTACTGTTCTGATCACAGGTGGTACTGGTGCTCTGGGCGCTCACGTTGC GCGCTGGGCCGCTGCTCAGGGTGTAGGTCACCTGCTGCTGGCTGGTCGTCGCGGTGCGGA TGCACCGGGCGTTGCAGAACTGCGTGCAGACCTGGAGGCATCTGGTGCAACTGTAACCGTA GCTTCTTGCGATGTGGCAGACCGCCGTGCTGTTCAGGAACTGCTGGATGGTATCCCGGATC GCCACCCGCTGACCGCCGTGTTCCATGCAGCTGGCGTCCTGGACGATGGTATGACGGCTGA TCTGGACACCGCACGTCTGGAGACCGTACTGGCCCCTAAAACCGAAGGTGCTCGTCTGCTG GACGAACTCACTCGTGATACCGAACTGAGCGCATTCGTACTGTTCTCTGGTTTCGCAGCAAC CGTGGGTTCTAGTGGTCAGGCCGGCTACGCGGCAGCTAACGCTCACCTGGACGCACTCGC

		GCAGCGCCGTCGTGCTGACGGTCTGCCGGCAACTTCTATTGCGTGGGTTCCGTGGACTGGC GGCGGTCTGGTTGACGAAGGTATCGAAGAGCGCCTGCTGCGTCGTGGTCTGAGTGCTATGG AACCGCGCCTGGCAGTGCTGGGTCTGTCTGCCCTGCTGAGCGGTGGCCCGGATGGTCCGG CTAACGTAGCACTGGTAGACGTTGATTGGCGTCGTTTCCTGCCGTCTTTCACCGCTTCCCGC CCGTCTCCGCTGCTGCGTGAGCTGCACCAGCCGTCTCCGGCGGACCCGGGTGCTACTGCT GGTCGTGACGGCACCCCAGGCGATGGCGCAGGTCGTACTCTGGCAGAACGT
	ACP	CTGCCGGGTCTGACCGACGCTGAACGTGACCGCCTGCTGCTTACAACTGTTCGCGCACAGG TGGCTTCCACTCTGGGTTATCCGGGCCCGGATGCTGTACCGCCGCGTCAGCAGTTCAAAGA ACTGGGTATCGACTCTCTGACTGCATTAGAACTGCGCAACGGTCTGCAAGCTGTTGCGGGC GTAGCACTGCCGGCTACTCTGGTTTTCGATCATCCGACCCCGGAGGCAGTTGCTGGTCACC TGTCCCGTACTCTGCTGGGTGATGGTGCACACACC
	linker	**AGCTCTTCCCTGTTTACCGAAATCGACCGCCTGGGCTCCGCACTGGGTTCCGCTGACCTGG ATGAAGCAGAACGTAACCTGGTAGCAACTCGTCTGCGTTCCCTGGCTTCTCGCTGGGATGC TGGTCCGTCTTCTGGCGCTGGTGACGTTGCACAGACTCTGGCTGACGCAGACCGCGCGGA GGTACTGGCATTCATCGATCGTGAACTGGGCATCTCTTCA**
PnD (Accession number AFJ05067)	**KS-AT**	ATGCGTTCCGAAGAACAGCTGCTGGACTACCTGAAACGTGTAACTGCTGACCTGCACGACAC CCGTGTGCGTCTGCAAGCTACCGAACAGGCCTCTCGCGAACCAATCGCAATCGTTGGCATG GGTTGTCGTTACCCAGGCGGTGTTTTCTCTCCGGAGGACCTGTGGCAGCTGGTTGCAGAAG GCCGCGACGTAATCGGTGACCTGCCGACTGATCGTTCTTGGAACCTGGACACTCTGTACGA TCCGGACCCGGCTGCACCGGGCACGACTTACGCTCGTGGTGGTGGTTTCGTTGCTACCGCG ACCCACTTCGACGCAGGCTTCTTCGGCATCTCCCCTCGTGAAGCAGCTGCAATGGACCCGC AGCAGCGTCTGGTACTGGAGACTGGTTGGGAAGCTCTGGAGCGCGCGGGCATCCCGCCGA CCTCTCTGGGTGGTTCTGATACTGGCGTTTACATCGGTACTGGTATGCAAGACCACATCATC CACCTGCAAAAAGCTCGCCAGGAGGCTGAAGGTTTCGTAGGCACGGGTAACGCGATCTCTG TGGTTTCTGGTCGTCTGGCTTACACCCTGGGTCTGGAAGGCCCGGCTGTTTCCGTAGACAC GGCCTGCTCTTCTTCTCTGGTTGCACTGCACCTGGCTGTTCAGGCTCTGCGTGCTGGTGAAT GCTCCCTGGCTCTGGCAGGGGGTGCTACGGTGATCAACACTCCGGTAATGTTCGTTGAGTT CTCTCGTCAGCGCGGTCTGTCCCCGGACGGTCGTTGCCGCTCTTTCGCTGCTGATGCTGAT GGCACGGGTTGGAGTGAAGGCGCGGGCGTACTGGTTCTGGAGAAACTGAGCGACGCTCGC CGTCACGGTCATCAGGTTCTGGCAGTTGTTCGTGGTTCCGCTGTTAACCAGGACGGTGCTTC TAACGGTCTGACTGCTCCGAACGGCCCGTCTCAGCAGCGCGTTATCCGTCGTGCTCTTACC GCAGCTGGCCTGACCGCAGATGAAGTCGACGTAGTAGAGGCACATGGTACTGGCACCCGTC TCGGGGACCCGATCGAAGCTCAGGCTCTGCTTGCAACCTACGGCCAGGGCCGTTCCGCTGA CCGTCCGCTGTGGCTGGGCTCTCTGAAGTCTAACATCGGTCACGCTCAGGCAGCAGCTGGC GTGGGTGGTGTTATCAAATCTGTAATGGCACTGCGTCACGGTGTTCTCCCGCAGACTCTGCA CGTTCACGCACCAACCCCGGAGGTAGACTGGTCTGCTGGCGAAGTGCGCCTGCTTACCGAG CGTCGTGACTGGCCACGTGCTGGTCGTCCGCGTCGTGCAGGTGTATCCAGCTTCGGTATCT CTGGCACTAACGCTCACCTGATCCTGGAGGAAGCACCAGATGAAGCGCCGGAAGAAGCTCC AGACGACGCAACCGACCGCGCACCGCGTTTGCGTGGTATGCCGTGGGTCGTGAGCGGTCG TGGTCGTCAGGCTCTGCGTGCGCAGGCTGCGCGTCTGCGTGACTGGGCAGCTGGTCACCC GGATGTTGCACAGGAAGAAGTGGGTCTGGCTCTCGCGACCCGCCGTTCTGTATTCGAGCAC ACCGCTGTGATCACCGGCGCAGACCATGATGAACTGCTGAACGGTTTATCTGCTCTGTCTGA AGATCGCCCGGCACCGGGTGTTCGTCTGGAGACAGCTGCTGGTGGTGGCCTCGCATTCGC GTTTACTGGTCAGGGTGCACAGCGTCCGGGTATGGGCCGTGGTCTGTATGAGACGTTCCCG GCTTACGCGGAAGCATTCGACGAGGCTTGCGCTGCACTGGACCCGCACCTGGAACGCCCG CTGGCTTCCGTTGTGTTTGCAGACGGCCCGGCTGACGCTGAGGCGCTGCAAACCACCGCTT ATGCCCAACCGGCTCTGTTCGCAGTTGAGACCGCTCTGTTCCACCTGATGCAGTCTATGGGT GTTCACCCGGACCTGCTGATCGGCCACTCTGTTGGCGAACTGAGCGCTGCGCACGCAGCTG GCGTTCTGAGCCTCCAAGATGCTGCACGCCTGGTAGCTGCTCGTGGTCGCCTCATGCAATC TCTGCCGGAGGACGGTCGTATGCTGGCCATCCAGGCATCTGAAGACGAGGTTCTTCCGTCC GTAGCTGAACTGAGCGAGAAGGCTGGTGGCGTTGCAGTTGCTGCTGTAAACGGCCCGGCAT CTGTAGTTGTATCTGGTCGTACTGAGGCTGTTCATACTCTCGAAAAAGAATTTGCAGGTCGC GGTCGTCGCGTACGTTACCTGGACGTGTCTCACGCATTCCACTCCCCACTGATGGACCCGG TTCTGGACGAGTTCGCACGCATCGCGGCTTCTGTTACCTTCCGTCCGGCTACCACCCCGGTT ATCTCTAACGTGACTGGTGACCTGATCGGTGATGACCGTCTGGCTGACCCGTCCTACTGGG CCGATCACATCCGTGCTACCGTTCGTTTCGCAGACGGTGTGCGTGCTCTGGCACGTGAACA GGTTGACACCGTAGTTGAACTGGGCCCGGACGCTGCACTGACCGCTCTTTGTGGTGAAATC CTGGACCAGGACACCGCAGCTTTCGTACCGACTCTGAGCCGTAAACACGACGAAACCTCTA CTTTCCTGACTGCTATGGCTCGACTGCACGCTCGTGGCATTCCGGTTCGTTGGCCGGCTGC GACCACTCCGTCTGTAGAAAGTGCTGATCTGCCGACCTACGCTTTCCAGCGTGAACGTCATT GGCTGGACGGTGTAGCGGAAACTGACGTGGCCGGCACTGGTCTGACC

	DH-KR	GGCATCGGTCACCCGCTGCTCCCGGCTGAAACTGCACTGCCGGGTACTGAGGGTGTTGTGC TGACTGGCTCTCTGAGTCTGCACGACCATGCTTGGCTGGCAGATCACGCTGTTCTGGGCGT AGTACTGGTGCCGGGAGCAGGTCTGCTGGACATGGCGCTGACAGCAGCTGAACATGCAGG CTGCACCCAGGTTGAAGAGCTGACCCTTGAATCCCCGCTGATTTTACCGGAGGTTGGTGCTC GTAGCGTACAGGTACTGGTAGGTGCGGCTCAGGACTCCGGGCAGCGTGCCATCACCATCCA CTCCCGTCCGCAGGACGGTGACCCGCACGCAGCATGGACCCGCCACGCCACCGGCCTGCT GGCTACTGGTCCACAGGAAGAACCGGCTACCTTCTCTGAAGCATGGCCGCCGACTGGCGCT GTACCAGTTCCGGTTGACGACCTGTACCTCCGCCTGACTGAAGGTGGTGTGGATTATGGTC CTTCTTTTCGTGGTCTGCGCGCAGCCTGGCGTCTGGGTGAAGACTTCTACGCAGACATCGA CCTGCCGCACCTGTCTGACGTAGAACGTTTCACCCTGCACCCGGCACTGTTAGACGCGGCT CTGCATTCTCTGGCACTGCCGGGTGCTATCCTCCATACCGGCCAGGCACACCTGCCGTTCT CTTGGTCTGGTGTGCGTCTGCACGCCTCTGGTGCAGACGCCCTGCGTATCCGCGTTCGTGG CACGGGTTCTTCCTCAGTATCTCTGGAACTGGCAGACGGTACTGGCCGCCCGGTAGCAACT GTTGGCGAGCTGGCTCTGCGTCCGGTATCTCAGGAACAACTGCGTACCCCGGGCGCAGATC CGACTAGCCTGTACACCGTAGAGTGGCCGGTAAAGGAACTGCCGGAACGTGCTGAAGGTTC TGCAACACGTGCGTGGGCAGTAATCGGTGCACCGGAGCCAGCTGGTTACGCGGTTGAAGGT GTTGAGCTGTCTCACTATGCTTCTCCGGCTACCCTGGCAGCAGCTCTGGATGCTACCGACGT GCCGGTTCCGGAGGCTGTTCTGGTGCCGTGCTTCTCTCCACACTCCCCGGGTTGTCCACCG GGTGCATCTGCTGGCCTGGTTGAGCGTGCGCACGACTATGCTGCAACTGTACTGGACCTGG TACGCACCTGGCTGGGTGATGCTCGTTTCGAGGCCTCTCGCCTGCTGCTGCTCACCCGTGG CGCAGTTACCCCAGACGCAGGCCACCACCCGCTGTCTGACGAGCTGGGCTCCCTGGCATG GGGTCTGGTCCGCACAGTACAGAACGAACACCCGGGTCGCGTCCTGGTAGCAGACCTGGAT GAAGATCCGGCTTCTTGGTCTGTTCTGCCCGGCGTGCTGGCACATGAGGAACCGCAGGTTG CTGTTCGTCGTGGTGTAGCGCACGTACCACGCCTGACTGCTGCACGCGCAGCAGCGGAAC GTGCAACCCCGTTCGACTCCGTGGGCACCGTACTGGTTACTGGGGGTACTGGTGGCCTGG GTTCTCTGCTGGCTCGCCACCTGGTAGTAGAGCACGGTGTACGCCATCTGCTGCTGACTTCT CGTCGCGGCCCGCAGGCTCCGGGTGCAGCTGCTCTGGCAGCTGAACTGACTGAACTGGGT GCACAGGTTACCGTTACCGCTTGCGACATGGCTGACTCCCCGGCGGTTGAAGAACTGCTGG GTAGCCTGCCGGCTGGCCACCCGCTGACCGCAGTAATCCACACCGCTGGTGTGCTGGATGA CGGTCTGGTTCAAGACCTGACCCCGGAACGTCTGAACACCGTGCTGCGCCCGAAAGCTGAT GCGGCTGTTGTTCTTGACCGCCTGACCCGTCACCTGGACCTGGCAGCTTTCGTCCTGTACTC TAGCGTAGGTGGCACCCTCGGTGGTCCGGGCCAGGGCAACTACGCCGCAGCTAACGCATT CCTGGACGCACTTGCTCAGCGTCGCCGCGCCGAGGGCCTGCCGGGTCTGTCTCTGGGTTG GGGTCTTTGGTCCGACACCACTGGTATGGCGGCTGAGATCGGTACTACCCATGTGGGCCGT CTGAACCGTTCTGGTCTGGTTACCATGTCTCCGGCTGAAGGTCTGGCACTGTTCGATGCTGC AATCTCCGGTGGCTATGGTCCGGTTGTCCTGCCTGTACGCCTGGACCTGCCGGCGCTGAAG GCACGCGCTACTTCTGGCGCTCTGCCGTCCGTTCTGGCAGACCTGGTACGTACCCCGGCGG GTCCGGCACGCGCGGCAAAATCTGGTACTGTGTCTGGTGCTGCTCTGCGTGCAGCT
	ACP	CTGTCTCAGCTGTCCGAAGACGAGCGCCGTCGTATGCTGCTCGACGTTGTACGTGAAAACG TTGCGGCTGTTTTGGGCCTGCGTCAGGACGGCGCTATGGACGAGGAGCAGCCGTTCAAAGA CCTGGGCCTGGACTCTCTGACGGCAGTAGAACTGCGCAACCGTCTGGCCAGCGCTACTGGC CTGCAACTGCCGGCAACCCTGCTGTTCGACCTGCCTTCTCCGAGCGCTCTGGCTCACCACC TGCTGACTCGCATGGGT**TCTCAAGAGAGTAAATCTCCGGTAGCTGAAGCAGTTGACCATCT GACCGCGCTGCTG**
	linker	**ACCACTCACGACGTAGGCGACCTGGAGCGTTCCCAGGTTACCGCCCGTCTGCGTTCTCTG CTGTGGCGCCTGGACGACGGTGGCGTAGAGTCTACTGAAGACGCGGCCGACCAGGCAGA AACTGACGACGACATCTTCGCACTGGTTGACCGTGAACTGGGTCTGGCTTCA**
DEBS thio-esterase (1MO2_A)		AGCGGGACTCCCGCCCGGGAAGCGAGCAGCGCTCTTCGCGACGGCTACCGGCAGGCGGG CGTGTCGGGCAGGGTCCGGTCCTACCTCGACCTGCTGGCGGGGCTGTCGGACTTCCGCGA GCACTTCGACGGCTCCGACGGGTTCTCCCTCGATCTCGTGGACATGGCCGACGGTCCCGGA GAGGTCACGGTGATCTGCTGCGCGGGAACGGCGGCGATCTCCGGTCCGCACGAGTTCACC CGGCTCGCCGGGGCGCTGCGCGGAATCGCTCCGGTTCGGGCCGTGCCCCAGCCCGGCTA CGAGGAGGGCGAACCTCTGCCGTCGTCGATGGCGGCGGTGGCGGCGGTGCAGGCCGATG CGGTCATCAGGACACAGGGGGACAAGCCGTTCGTGGTGGCCGGTCACTCCGCGGGGGCAC TGATGGCCTACGCGCTGGCGACCGAACTGCTCGATCGCGGGCACCCGCCACGCGGTGTCG TCCTGATCGACGTCTACCCGCCCGGTCACCAGGACGCGATGAACGCCTGGCTGGAGGAGC TGACCGCCACGCTGTTCGACCGCGAGACGGTGCGGATGGACGACACCAGGCTCACCGCCC TGGGCGCCTACGACCGCCTCACCGGTCAGTGGCGACCCCGGGAAACCGGGCTGCCGACGC TGCTGGTCAGCGCCGGCGAGCCGATGGGTCCGTGGCCCGACGACAGCTGGAAGCCGACGT GGCCCTTCGAGCACGACACCGTCGCCGTCCCCGGCGACCACTTCACGATGGTGCAGGAAC

		ACGCCGACGCGATCGCGCGGCACATCGACGCCTGGCTGGGCGGAGGGAATTCGAGCTCCG TCGAC
Npt (OSY40025) Codon optimized		ATGATTGAGAAGTTACTCCCGGCGCCAGTCAGAACGGCAGAGACTTTCGACGATGCGCCTTT ATCTGAAATGTTCCCCGAAGAGTGGGCGCAGGTTGCAAACGCTGTACCCAAACGCCAACGT GAGTTCGGTACTGTACGAGGGTGCGCTCGTCGTGCCCTGGCCGAGCTTGGCTTCGCTCCG GCACCATTGCTGCCTGGACCTCATCGTGAGCCGCAGTGGCCAGATGGGGTTGTGGGCGCG ATGACGCACTGCGCGGGATATCGCGCTGTAGCGGTGGCACGCGCCGGCGAAGTTCGCACA ATCGGCCTGGATGCCGAACCGAATCTCCCACTAAATGACCCGGGCGTTCTTGACCTGGTGA CATTACCGGAAGAACGGGACCAGATCCGGCGCCTCGCCGCCCTTCAACCGGAAGTCTGTTG GGATCGCTTGGTCTTTTCCGCAAAAGAAAGTGTCTACAAAGCCTGGTTTCCGCTGACGCGCC GTTGGTTGGATTTTGAAGAAGCACTGCTGACCTTTGATCCGACCAACGCGACCTTTACCGCG CAGCTGCTGGTGCCGGGCCCGGTGGTTGATGGTCGTGAACTGACCGAATTTTCGGGTCGTT GGCTGGTGGGTAGCGGTCTGGTCGTTACCGCGATTGTGGAAATGGTGTCA

References

1. Chen, Y. L.; Zhao, J.; Liu, W.; Gao, J. F.; Tao, L. M.; Pan, H. X.; Tang, G. L., Identification of phoslactomycin biosynthetic gene clusters from Streptomyces platensis SAM-0654 and characterization of PnR1 and PnR2 as positive transcriptional regulators. *Gene* **2012,** *509* (2), 195-200.
2. Vagstad, A. L.; Bumpus, S. B.; Belecki, K.; Kelleher, N. L.; Townsend, C. A., Interrogation of global active site occupancy of a fungal iterative polyketide synthase reveals strategies for maintaining biosynthetic fidelity. *J Am Chem Soc* **2012,** *134* (15), 6865-77.
3. Koryakina, I.; Kasey, C.; McArthur, J. B.; Lowell, A. N.; Chemler, J. A.; Li, S.; Hansen, D. A., Inversion of Extender Unit Selectivity in the Erythromycin Polyketide Synthase by Acyltransferase Domain Engineering. **2017,** *12* (1), 114-123.
4. Kushnir, S.; Sundermann, U.; Yahiaoui, S.; Brockmeyer, A.; Janning, P.; Schulz, F., Minimally Invasive Mutagenesis Gives Rise to a Biosynthetic Polyketide Library. *Angewandte Chemie International Edition* **2012,** *51* (42), 10664-10669.
5. Del Vecchio, F.; Petkovic, H.; Kendrew, S. G.; Low, L.; Wilkinson, B.; Lill, R.; Cortes, J.; Rudd, B. A.; Staunton, J.; Leadlay, P. F., Active-site residue, domain and module swaps in modular polyketide synthases. *J Ind Microbiol Biotechnol* **2003,** *30* (8), 489-94.
6. Vogeli, B.; Geyer, K.; Gerlinger, P. D.; Benkstein, S.; Cortina, N. S.; Erb, T. J., Combining Promiscuous Acyl-CoA Oxidase and Enoyl-CoA Carboxylase/Reductases for Atypical Polyketide Extender Unit Biosynthesis. *Cell Chem Biol* **2018,** *25* (7), 833-839.e4.
7. Sundermann, U.; Bravo-Rodriguez, K.; Klopries, S.; Kushnir, S.; Gomez, H.; Sanchez-Garcia, E.; Schulz, F., Enzyme-directed mutasynthesis: a combined experimental and theoretical approach to substrate recognition of a polyketide synthase. *ACS chemical biology* **2013,** *8* (2), 443-50.

Chapter III

Function of the type II thioesterase associated with the phoslactomycin polyketide synthase

This chapter is written in manuscript style and in preparation for submission.

Function of the type II thioesterase associated with the phoslactomycin polyketide synthase

Kyra Geyer [a], Steffen Hartmann [a], Randolph R. Singh [b], Tobias J. Erb [a], [c]

[a]Department of Biochemistry and Synthetic Metabolism; Max Planck Institute for Terrestrial Microbiology, Karl-von-Frisch-Str. 10, D-35043 Marburg, Germany, toerb@mpi-marburg.mpg.de [b]Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, avenue du Swing 6, L-4367 Belvaux, Luxembourg [c]LOEWE Center for Synthetic Microbiology (Synmikro), Karl-von-Frisch-Str. 16, D-35043 Marburg, Germany.

Author contributions: K.G and T.J.E conceived the project. K.G designed the experiments, K.G and S.H performed experiments and analyzed the data. R.S evaluated the tandem mass spectrometric data. K.G, and T.J.E wrote the manuscript with contributions from all authors.

Abstract

Polyketide synthases (PKS) join basic building blocks in an assembly line fashion, which leads to the production of structural highly diverse polyketides. Most PKS biosynthetic gene clusters contain a gene encoding for a type II thioesterase (TEII). TEIIs restore the productivity of PKS by removing aberrant modifications on the acyl-carrier proteins (ACP), which cause a blockage of the assembly line. This function is commonly referred to as an editing function. Here, we investigated the function of PnG, the TEII from the phoslactomycin (Pn PKS) gene cluster, on its cognate modular type I PKS *in vitro*. Detailed biochemical analysis revealed a strong preference of PnG towards alkyl-ACPs over (alkyl)malonyl-ACPs, supporting the editing role of this TEII. In phoslactomycin pentaketide assays, we showed the PnG dependent increase in production of native as well as non-native polyketides. We additionally demonstrate that PnG is able to functionally replace the terminal TEI of the chimeric Pn PKS system *in vitro* by releasing terminal polyketides from the assembly line. Furthermore, we provide insights in the origin of the aberrant assembly line-blocking modifications.

Introduction

Polyketide synthases (PKS) are multifunctional enzyme complexes that biosynthesize natural products with great structural complexity and diverse bioactivity. The structural backbone of these compounds is synthesized through condensation of simple (alkyl)malonyl-CoA precursors via the successive function of discrete PKS domains in an assembly line fashion. The final polyketide backbone is usually released by a covalently bound terminal thioesterase domain. Besides the terminal thioesterase, most PKS, but also nonribosomal peptide synthethase (NRPS) gene clusters, encode a type II thioesterase (TEII). The majority of the information on TEIIs was gained by gene disruption and replacement strategies, which showed effects from complete loss of product formation, over reduced amounts, to no change in natural product titers [1-9]. Co-expression of TEIIs was successfully used to increase the yield of natural product production [10-12], while strong overexpression resulted in reduced product amounts, indicating a tradeoff between removing blocking modifications and productive intermediates [6, 13].

It was suggested that TEIIs display an editing function, whereby aberrant precursors that block the assembly line are removed, to allow continued synthesis [4, 14]. Other reported functions of TEIIs are the release of the final product [12, 15, 16], the control of the pool size of CoA-thioester precursors [1], the control of starter units [17, 18], and the unspecific release of short acyl-residues and late stage intermediates [19]. The blockage of PKS assembly lines has been proposed to originate from a slow, unproductive decarboxylation of (alkyl)malonyl extender units bounds to their cognate acyl-carrier protein (ACP) [20-24]. Another potential source of unreactive alkyl-ACPs in the PKS assembly line is the conversion of *apo*-ACPs into their active *holo* forms through the attachment of the phosphopantetheine prosthetic (4'PP) group to the ACP by specific transferases [10, 25, 26]. The role of TEIIs in editing (i.e. clearing) unreactive alkyl-units is supported by data that shows that these enzymes prefer decarboxylated (i.e. alkyl-) over carboxylated residues, and ACP tethered over CoA/*N*-acetylcysteamine tethered substrates [27].

Two models have been proposed for the general editing function of TEIIs. In the high specificity model, the TEII efficiently removes aberrant acyl units. In the low

specificity model, the TEII removes both aberrant and correct units with low efficiencies [20, 27]. Since correct substrates will be efficiently processed by the PKS assembly line, they are protected from the TEII, while incorrect, aberrant units will lead to assembly line stalling and become exposed to the TEII, followed by removal.

Only a few TEIIs have been studied *in vitro* together with the cognate NRPS or PKS system. YbtT, the TEII from yersiniabactin hybrid NRPS/PKS did not have an effect on the *in vitro* system, when only the native substrates were present. Later it was found to remove non-canonical precursors that blocked the synthesis *in vitro*, therefore restoring full biosynthetic activity [13, 28]. In the reconstituted type II PKS of enterocin and tetracenomycin, the TEII EncL and ZhuC, respectively, were shown to impact starter unit selection [17, 18]. Besides selection of starter units, regio-specific selection and incorporation of extender units must be assured. This becomes especially apparent in modular type I PKS, that incorporate multiple extender units, such as the phoslactomycin (Pn) PKS which uses cyclohexanecarboxyl-CoA as starter unit and malonyl- and ethylmalonyl-CoA as extender units. Here we investigated the role of PnG, the TEII from the phoslactomycin biosynthetic gene cluster [29], on the production of phoslactomycin derivatives using a recently established *in vitro* system [30]. We biochemically characterized PnG, and provide data that support both decarboxylation during polyketide assembly and miss-loading during protein production as the origin of the aberrant units, blocking the assembly line. PnG exhibits a broad substrate tolerance and is capable of releasing final polyketides, replacing the function of the terminal thioesterase while at the same time, increasing the *in vitro* polyketide production of native and non-native polyketides with altered site chains.

Results and discussion

Heterologous expression and purification of PnG and Pn PKS enzymes.

6xHis-tagged PnG protein was purified to homogeneity by Ni-NTA purification, followed by size exclusion chromatography (Figure S1). PnG eluted as a monomeric protein with 34.4 kDa size (calculated 30.17 kDa), which is in line with the general assumption that TEIIs are functional as monomers [27, 31, 32]. Active site knockout PnG S93G showed the same elution patterns as the wild type. Enzymes of the Pn PKS were purified as previously described [30].

PnG hydrolyses decarboxylated acyl-ACPs.

First, we determined the kinetic parameters of PnG. For this, we used two different ACPs from the native phoslactomycin PKS (Pn PKS), ACP_{LD} (from the loading domain of PnA) and ACP2 (from the first module of PnB) as substrate templates (Table 1, Figure S2). We loaded these ACPs with different acyl-residues using Npt, the promiscuous phosphopantetheinyl-transferase from *Streptomyces platensis* [30]. PnG was equally active with both ACPs, which is in line with the observation that TEIIs generally show low specificity towards the carrier protein [2, 6, 13].

Table 1: Reaction parameters for PnG hydrolytic activity with various acyl-residues bound to ACP_{LD} and ACP2 with the standard error of the mean.

Substrate	k_{cat} [min^{-1}]	K_M [µM]	k_{cat}/K_M [$M^{-1}s^{-1}$]
decarboxylated substrates			
Acetyl-ACP_{LD}	618 ± 45	47 ± 17	$2.2x10^5 \pm 8x10^4$
Butyl- ACP_{LD}	493 ± 38	23 ± 9	$3.6x10^5 \pm 1.4x10^5$
Acetyl-ACP2	965 ± 56	27 ± 7	$6x10^5 \pm 1.6x10^5$
Propionyl-ACP2	2050 ± 358	139 ± 57	$2.5x10^5 \pm 1.1x10^5$
Butyl-ACP2	353 ± 39	23 ± 12	$2.6x10^5 \pm 1.4x10^5$
carboxylated substrates			
Malonyl- ACP_{LD}	259 ± 31	293 ± 109	$1.5x10^4 \pm 5.8x10^3$
Ethylmalonyl- ACP_{LD}	10 ± 0.8	501 ± 113	$3.3x10^2 \pm 8x10^1$
Malonyl-ACP2	77 ± 13	164 ± 99	$7.8x10^3 \pm 4.9x10^3$
Methylmalonyl-ACP2	20 ± 5	369 ± 196	$9x10^2 \pm 5.3x10^2$
Ethylmalonyl-ACP2	2 ± 0.3	168 ± 72	$2x10^2 \pm 9x10^1$

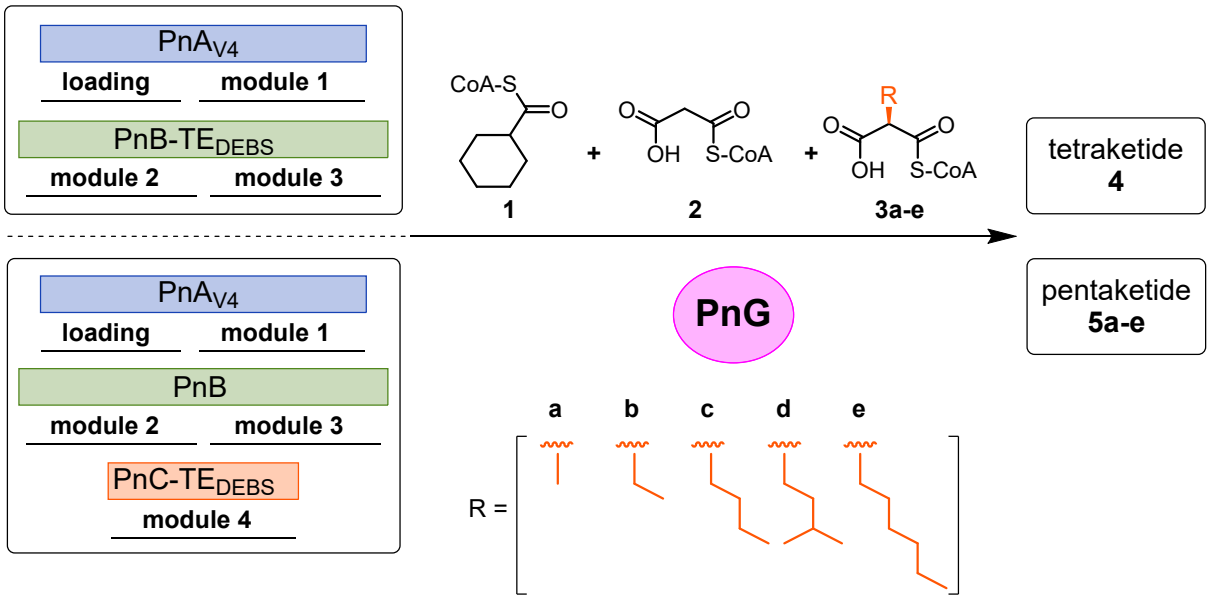
Our kinetic analysis showed distinct differences in the hydrolytic activity of PnG between the two substrate classes tested (Table 1). PnG hydrolyzes all tested alkyl-ACPs representing the decarboxylated substrates (acetyl-, propionyl- and butyl-ACP) with similar specificity constants ranging from $2.2x10^5$ to $6x10^5$ M^{-1} s^{-1}, indicating no preference for chain length as for instance observed for RedJ, the TEII in the prodiginine biosynthesis cluster [31]. When we tested carboxylated (alkyl)malonyl-ACP substrates, however, we observed a 45 to 50-fold preference for malonyl- over ethylmalonyl-ACP indicating that PnG is able to distinguish between shorter and longer carboxylated substrates, which might protect ethylmalonyl-ACP from futile hydrolysis.

When comparing hydrolytic activity between (alkyl)malonyl-ACP substrates and their decarboxylated analogs, we observed that PnG displays a strong discrimination towards the latter. PnG preferred acetyl-ACP over malonyl-ACP with a 14 to 75 fold increased specificity constant. In the case of propionyl-/methylmalonyl-ACP and butyl-/ethylmalonyl-ACP we determined a 150-fold and 1000- to 1500-fold increase in specificity, respectively. Preferences for decarboxylated over carboxylated substrates were also reported with other TEII [5, 13, 27, 31], however not as pronounced as in the case of PnG. Our kinetic data shows that this preference for the decarboxylated substrate is caused by both high K_M values, as well as low turnover rates. Altogether, these results suggest that the hydrolysis of carboxylated substrates from ACPs is negligible under physiological conditions.

PnG leads to higher production of polyketides *in vitro*.

To investigate if PnG, encoded in the biosynthetic gene cluster of phoslactomycin [29] has an effect on phoslactomycin biosynthesis, we tested the enzyme in the context of the recently established Pn PKS *in vitro* system for the production of tetra- and pentaketides (Scheme1) [30]. In this system, polyketide biosynthesis is initiated by PnA_{V4} with cyclohexanecarboxyl-CoA and malonyl-CoA and continued by PnB using two malonyl-CoA, yielding tetraketide (**4**). For production of pentaketide (**5**), another extension of **4** is followed through PnC. While PnC incorporates ethylmalonyl-CoA (**3b**) in the native context (Figure S3), it was shown that its acyltransferase (AT) is promiscuous and

accepts different non-native extender units (**3a**, **3c-3e**) with a preference for medium side chain lengths substrates **3b-3d**, which leads to pentaketides **5a-5d.**



Scheme 1: Assay scheme for characterization of PnG. Tetraketide (**4**) or pentaketide (**5**) production is initiated by PnA$_{V4}$, terminates with PnB-TE$_{DEBS}$ in case of **4**, or with PnC-TE$_{DEBS}$ in case of **5.** Substrates used are cyclohexanecarboxyl-CoA (**1**), malonyl-CoA (**2**) and in single extender unit pentaketide production assays one, in the case of competitive assays all five *α*-substituted extender units (**a** methyl-, **b** ethyl-, **c** butyl-, **d** 3-methylbutyl- and **e** hexyl-malonyl-CoA). Ethylmalonyl-CoA **3b** is the native substrate of PnC, however **3a** to **3e** are incorporated into the pentaketide.

In the tetraketide system, the presence of PnG increased production of **4** between four- and five-fold, independent of the PnG concentration (Figure 1A, Figure S4). Addition of PnG S93G, a catalytic knockout control, did not impact polyketide biosynthesis (Figure S5). To test whether the mis-loading of ACP$_{LD}$ might affect the assembly line we tested the effect of acetyl-CoA, a common metabolite, on tetraketide production. When adding 1 mM acetyl-CoA to the tetraketide assays, production of **4** was unaffected by the presence or absence of PnG, which ruled out acetyl-CoA mis-loading of ACP$_{LD}$ as a significant mechanism for assembly line stalling (Figure 1A). This was further supported by the fact that no production of modified **4**, as result of acetate starter incorporation was observed in presence or absence of PnG. Altogether, these experiments demonstrate a positive effect of PnG onto the biosynthesis of **4**.

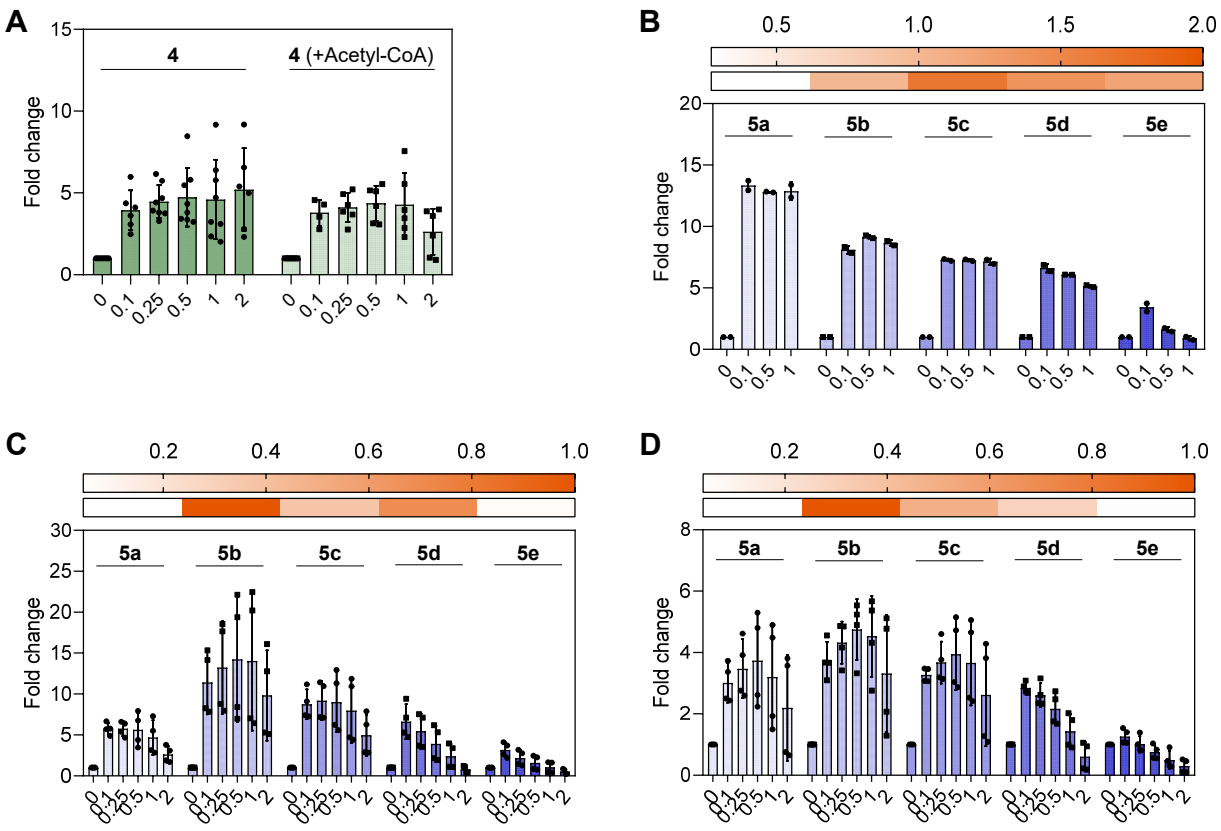


Figure 1: Effect of increasing PnG concentration on polyketide production *in vitro.* Concentrations of PnG ranged between 0.1 and 2 molar equivalents of ACPs present in the assay, indicated on the x-axis. All values were set relative to the control without PnG. Samples were taken after 2 h and measured by mass spectrometry. **(A)** Tetraketide **4** production with the native substrates and additionally with acetyl-CoA. **(B)** Pentaketide **5a-e** production with **2** and one *α*-substituted malonyl-CoA derivative **3a-e** as extending substrates. **(C)** Pentaketide production in competitive conditions with all substrates **1**, **2** and **3a-e**. (**D**) Pentaketide **5a-e** production in competitive conditions with pretreated Pn PKS enzymes, cleared of potential ACP-tethered acyl-residues. Pre-incubation with PnG assures removal of all acyl-residues. In **B**, **C** and **D** the heat map indicates the initial product distribution, set relative to the native product **5b**. Residues indicated are: **a**=methyl-, **b**=ethyl-, **c**=butyl-, **d**=3-methlbutyl-, **e**=hexyl-residues. A negative control with PnG S93G catalytic knockout was used (Figure S5). Data points are from multiple biological (**A**, **C**, **D**) and technical (**A**, **B**, **C**, **D**) replicates. For a list of the product masses refer to Table S1.

For the pentaketide system we supplied **1**, **2** and either one *α*-substituted extender unit alone, or all five *α*-substituted extender units at the same time in a competitive assay (Scheme 1). This aimed to detect an effect of PnG on Pn PKS biosynthesis in the

presence of native (**2** and **3b**) or non-native (**3a**, **3c**-**3e**) extender units and to detect an effect of PnG on Pn PKS system that is challenged with multiple substrates simultaneously.

In pentaketide assays, we observed incorporation of the respective extender unit in each case (Figure 1B-D, Figure S4). In the absence of PnG, we detected an initial pentaketide distribution between **5a**-**5e** as reported before, with minor amounts of **5a** and with **5b-d** as the main products [30]. Tandem mass spectrometry confirmed the structures of pentaketides **5a**-**5e** (Figure S6). As for the tetraketide system, addition of PnG in the pentaketide system strongly increased product formation. In assays with only one single *α*-substituted extend present (Figure 1B), formation of **5a** was increased 13-fold (note that initial amount had been close to detection limit), while **5b**-**d** were increased around nine-fold. Finally, production of **5e** was changed only three-fold and remained unaltered upon high concentrations of PnG.

In competitive extender unit assays, containing all five *α*-substituted extender units (Figure 1C, Figure S4) we observed an increase on the production of **5a**-**5e**, particularly upon low concentrations of PnG. The beneficial effect of PnG was most pronounced on **5b**, the native product, with a 14-fold increase, followed by **5c** (nine-fold), **5d** (seven-fold), **5a** (six-fold) and **5e** (three-fold). Production of **5a** was also increased several-fold compared to the initial concentrations that had been close to the detection limit. Overall, these results showed that PnG generally has a beneficial effect on tetra- and pentaketide production *in vitro* and increases of up to 14-fold more product can be observed.

As mentioned above, TEIIs remove unwanted modifications on the reactive thiol group of the 4'-PP cofactor of ACPs, that block the assembly line. These modifications can originate from the slow, unproductive decarboxylation of (alkyl)malonyl-ACPs or the mis-loading of ACPs by promiscuous phosphopantetheinyl-transferases. To test if mis-loading of ACPs during protein production in *E. coli* BAP1 plays a role, we pre-treated purified Pn PKS enzymes with Npt, the native transferase from *S. platensis*, CoA and PnG to ensure complete activation and clear all acyl-residues from ACPs. By size exclusion chromatography PnG and Npt were removed and Npt was omitted from the assays to circumvent mis-loading of any residual *apo*-ACPs. Notably, with these pretreated enzymes, increase of pentaketide production was still observed (Figure 1D,

Figure S4). As before, low to medium concentrations of PnG showed the most prominent effect. Production of **5a**, **5b**, **5c** and **5d**, was increased between three- and five-fold, while the effect on **5e** was negligible. Overall, these experiments suggest that the action of PnG is a cumulative effect of clearing mis-loaded ACPs before and during Pn biosynthesis, where slow, unproductive decarboxylation of (alkyl)malonyl-ACPs takes place.

We were interested if PnG impacts all Pn PKS proteins similarly. Pn polyketide biosynthesis can be initiated by PnA_{V4} or by PnB, when providing the diketide SNAC substrate analog and terminated with PnB-TE_{DEBS} (for **4**) or PnC-TE_{DEBS} (for **5**) [30]. Production of **4** and **5** when initiated with PnA_{V4} was increased approx. five- and 14-fold, respectively, upon addition of PnG (Figure 1). When bypassing PnA_{V4} with the diketide SNAC ester no beneficial effect on the production of **4** was observed, while an eight-fold increase can be detected for production of **5** (Figure S7A). This data indicates either a stronger impact of PnG on PnA_{V4} and PnC-TE_{DEBS} than on PnB or an inhibition of PnG caused by the presence of the diketide SNAC. An inhibition of PnG would lead to a less pronounced effect when adding the SNAC ester to the full pentaketide **5** assembly line. To investigate this further, we supplied the SNAC thioester to pentaketide assays initiated with PnA_{V4} and PnB as a control and observed a decrease of PnGs production promoting effect (Figure S7B). This suggests that PnG binds the SNAC thioester as substrate and is occupied by it (due to slower turnover), showing a negative effect of the presence of these polyketide analogs on this TEII.

In general, low amounts of PnG resulted in a maximum increase of product formation. High PnG concentrations on the other hand had a less beneficial or even negative impact on biosynthesis, particularly in case of non-native, long *α*-substituents (i.e. hexyl-residue of **3e**). This observed negative effect could be caused by increased PnG catalyzed clearing of non-native (alkyl)malonyl extender units bound to the ACP or the corresponding CoA esters, as some TEIIs were shown to hydrolyze CoA thioesters at low activity [26, 27]. To rule out the latter, we incubated PnG with all CoA thioester substrates used in this study. Even after several hours of incubation, CoA ester hydrolysis was below the detection limit (Figure S8). We concluded that the negative effect of increased PnG concentrations on the production of non-native pentaketides was caused by an increased hydrolysis of ACP bound (alkyl)malonyl extenders, probably due their

slower turnover and thus increased exposition to PnG. Notably, this effect of PnG leads to a shaping of the product profile in competitive pentaketide assays towards the native product (**5b**) (Figure S9) and supports the low specificity model of TEIIs, according to which hydrolysis of native extender units is rather inefficient and does not compete with chain elongation [20, 27].

Release of polyketides by PnG.

Finally, we assessed whether PnG is able to release polyketides from the assembly line. Therefore, we first tested the pentaketide assembly line in a competitive assay without the terminal thioesterase TE_{DEBS} (PnA_{V4}+PnB+PnC) (Figure 2, Figure S10). Without TE_{DEBS}, the initial product distribution pattern was similar to assays terminating with TE_{DEBS}, while the product amount was strongly reduced below 5%. This supports the functionality of TE_{DEBS} in that system. Addition of PnG in low to medium concentrations increased product formation for all pentaketides. Production of **5a** and **5e** was increased around one order of magnitude, while formation of **5c** and **5d** was increased around two orders of magnitude and formation of the native product **5b** by 200-fold. Compared to the system terminating with TE_{DEBS}, final product levels reached 50% (**5d**), 70% (**5c**) and 90% (**5b**) with PnG, respectively.

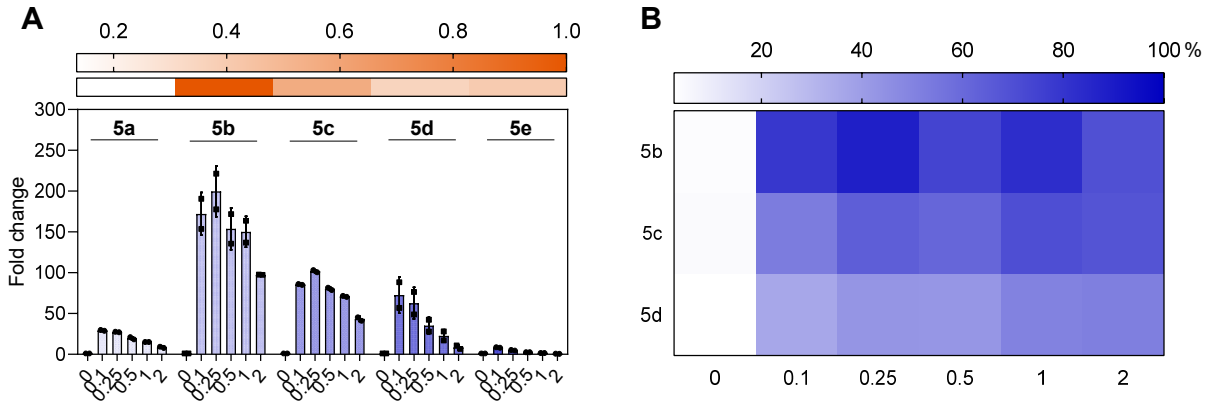


Figure 2: PnG releases pentaketides. Pentaketide production in competitive assays lacking a terminal thioesterase. Concentrations of PnG ranged between 0.1 and 2 molar equivalents of ACPs present in the assay, indicated on the x-axis. All samples were taken after 2 h incubation and analyzed by mass spectrometry. (**A**) The heat map displays the initial relative amounts of the pentaketides **5a**-**e** in the absence of PnG, set relative to the native pentaketide **5b**. A strong increase of product formation upon addition of

PnG can be observed for all pentaketides, in particular **5b-5d**. (**B**) Heat map showing the production yield in percentage of assays terminating without (PnA_{V4}, PnB, PnC) compared to assays terminating with TE_{DEBS} (PnA_{V4}, PnB, PnC-TE_{DEBS}) with increasing PnG concentrations. Results shown for the most prominent products **5b**, **5c** and **5d**. Residues indicated are: **a**=methyl-, **b**=ethyl-, **c**=butyl-, **d**=3-methlbutyl-, **e**=hexyl-residues.

We then tested the hexaketide system of phoslactomycin [30] whether a similar effect with PnG can be observed. Unexpectedly and in contrary to what we observed in the pentaketide system, the detected hexaketide amount was increased two-fold in the absence of TE_{DEBS}, pointing towards incapability of this terminal TE for hexaketide release. Addition of PnG to the system with TE_{DEBS} resulted in five-fold, without TE_{DEBS} in 20-fold product increase (Figure S11). Overall this data demonstrates that PnG is capable of releasing products from the terminal modules of Pn polyketide systems. Furthermore, it shows that PnG can functionally replace the chimeric TE_{DEBS} in the pentaketide system and circumvent product release limitations caused by the terminal TE due to its broad substrate spectrum.

We were further interested whether PnG was not only able to release final products, but also premature products (i.e. , di-, tri- and tetraketide) in pentaketide assays. The growing polyketide is intermolecular translocated between PnA_{V4} and the first module of PnB, and also between the second module of PnB and PnC. In case of inefficient module interaction and/or translocation downstream, the polyketide chain will be exposed on the ACP and thus become a potential substrate for PnG. Consequently, an accumulation of diketide and tetraketide in the pentaketide system should be observed. Indeed, PnG dependent accumulation of di- and particularly tetraketide can be observed while no accumulation of triketide (intramolecular transfer) can be detected (Figure S12). This supports that PnG is able to release ACP tethered polyketides during translocation in an *in vitro* setting.

Final discussion

We biochemically characterized the type II thioesterase PnG and studied the effects of this enzyme on its cognate phoslactomycin modular type I PKS *in vitro*. PnG shows a strongly pronounced preference for decarboxylated acyl-residues bound to ACPs

compared to their carboxylated counterparts with up to 1500-fold increased specificity constant (k_{cat}/K_M). In addition, this enzyme prefers smaller, carboxylated substrates like malonyl-ACP over ethylmalonyl-ACP (50-fold higher k_{cat}/K_M). The data points towards that PnG removes aberrant ACP-tethered acyl-residues.

Using the recently established *in vitro* Pn PKS system [30], we show that PnG leads to a higher polyketide production in the presence of the native but also non-native substrates. Addition of PnG to extender unit competition assays allows increases of up to 14-fold higher product amounts. Our experiments demonstrate that the phoslactomycin assembly line is partially blocked by mis-loading during expression. This highlights the importance of editing TEIIs, as 4'phosphopantetheinyl transferases generally have a broad substrate tolerance and are able to transfer acylated phosphopantetheine arms to ACP [30, 33-35]. Additionally, the assembly line stalls during biosynthesis, most likely due to a slow, unwanted decarboxylation of ACP-tethered (alkyl)malonyl residues. This decarboxylation activity has been conferred to the ketosynthase domains, leaving an ACP-tethered alkyl-residue that is unreactive for the subsequent condensation step [36-38].

Notably, PnG does not only show hydrolysis activity towards short acyl-chains but also towards polyketides, as indicated by both the release of intermediates and in particular, the efficient penta- and hexaketide release of the system lacking a terminal thioesterase. This is simultaneously accompanied by the increase of product formation due to the editing role of PnG. In our *in vitro* system, PnG (and potentially other TEII) can therefore be used to replace the terminal TE, which might circumvent artifacts caused by their specificities, as seen for the hexaketide production.

In substrate competition assays PnG shapes the product spectrum towards the native product. This effect becomes especially visible with increasing PnG concentrations, as the probability of removing modifications that are bound to the ACPs with decreased processing-efficiency increases. In summary, the low specificity towards decarboxylated substrates, ACPs and the shaping of the product profile support the low specificity model of TEIIs for PnG. The activity of PnG, given the fine balancing of its optimal concentration, overall increases product formation, even for non-native products, which makes it interesting for the biosynthesis of modified polyketides.

Acknowledgements

We thank Dr. N. Paczia for running the LC-MS samples and Dr. U. Linne for hosting us in his laboratory for the mass spectrometric measurements of the ACP domains. This work was supported by the Max Planck Society and the LOEWE program (Landes-Offensive zur Entwicklung wissenschaftlich-ökonomischer Exzellenz) of the state of Hessen within the framework of the MegaSyn Research Cluster.

References

1. Wu, H.; Liang, J.; Gou, L.; Wu, Q.; Liang, W. J.; Zhou, X.; Bruce, I. J.; Deng, Z.; Wang, Z., Recycling of Overactivated Acyls by a Type II Thioesterase during Calcimycin Biosynthesis in Streptomyces chartreusis NRRL 3882. *Applied and environmental microbiology* **2018,** *84* (12).
2. Kotowska, M.; Pawlik, K.; Butler, A. R.; Cundliffe, E.; Takano, E.; Kuczek, K., Type II thioesterase from Streptomyces coelicolor A3(2). *Microbiology (Reading, England)* **2002,** *148* (Pt 6), 1777-1783.
3. Doi-Katayama, Y.; Yoon, Y. J.; Choi, C. Y.; Yu, T. W.; Floss, H. G.; Hutchinson, C. R., Thioesterases and the premature termination of polyketide chain elongation in rifamycin B biosynthesis by Amycolatopsis mediterranei S699. *The Journal of antibiotics* **2000,** *53* (5), 484-95.
4. Butler, A. R.; Bate, N.; Cundliffe, E., Impact of thioesterase activity on tylosin biosynthesis in Streptomyces fradiae. *Chemistry & biology* **1999,** *6* (5), 287-92.
5. Zhou, Y.; Meng, Q.; You, D.; Li, J.; Chen, S.; Ding, D.; Zhou, X.; Zhou, H.; Bai, L.; Deng, Z., Selective removal of aberrant extender units by a type II thioesterase for efficient FR-008/candicidin biosynthesis in Streptomyces sp. strain FR-008. *Applied and environmental microbiology* **2008,** *74* (23), 7235-42.
6. Kim, B. S.; Cropp, T. A.; Beck, B. J.; Sherman, D. H.; Reynolds, K. A., Biochemical evidence for an editing role of thioesterase II in the biosynthesis of the polyketide pikromycin. *The Journal of biological chemistry* **2002,** *277* (50), 48028-34.
7. Kotowska, M.; Ciekot, J.; Pawlik, K., Type II thioesterase ScoT is required for coelimycin production by the modular polyketide synthase Cpk of Streptomyces coelicolor A3(2). *Acta biochimica Polonica* **2014,** *61* (1), 141-7.
8. Chen, S.; Roberts, J. B.; Xue, Y.; Sherman, D. H.; Reynolds, K. A., The Streptomyces venezuelae pikAV gene contains a transcription unit essential for expression of enzymes involved in glycosylation of narbonolide and 10-deoxymethynolide. *Gene* **2001,** *263* (1-2), 255-64.
9. Xue, Y.; Zhao, L.; Liu, H. W.; Sherman, D. H., A gene cluster for macrolide antibiotic biosynthesis in Streptomyces venezuelae: architecture of metabolic diversity. *Proceedings of the National Academy of Sciences of the United States of America* **1998,** *95* (21), 12111-6.
10. Kotowska, M.; Pawlik, K., Roles of type II thioesterases and their application for secondary metabolite yield improvement. *Applied microbiology and biotechnology* **2014,** *98* (18), 7735-46.

11. Pfeifer, B.; Hu, Z.; Licari, P.; Khosla, C., Process and metabolic strategies for improved production of Escherichia coli-derived 6-deoxyerythronolide B. *Applied and environmental microbiology* **2002,** *68* (7), 3287-92.
12. Hua, K.; Liu, X.; Zhao, Y.; Gao, Y.; Pan, L.; Zhang, H.; Deng, Z.; Jiang, M., Offloading Role of a Discrete Thioesterase in Type II Polyketide Biosynthesis. *mBio* **2020,** *11* (5).
13. Ohlemacher, S. I.; Xu, Y.; Kober, D. L.; Malik, M.; Nix, J. C.; Brett, T. J.; Henderson, J. P., YbtT is a low-specificity type II thioesterase that maintains production of the metallophore yersiniabactin in pathogenic enterobacteria. *The Journal of biological chemistry* **2018,** *293* (51), 19572-19585.
14. Schneider, A.; Marahiel, M. A., Genetic evidence for a role of thioesterase domains, integrated in or associated with peptide synthetases, in non-ribosomal peptide biosynthesis in Bacillus subtilis. *Archives of microbiology* **1998,** *169* (5), 404-10.
15. Harvey, B. M.; Hong, H.; Jones, M. A.; Hughes-Thomas, Z. A.; Goss, R. M.; Heathcote, M. L.; Bolanos-Garcia, V. M.; Kroutil, W.; Staunton, J.; Leadlay, P. F.; Spencer, J. B., Evidence that a novel thioesterase is responsible for polyketide chain release during biosynthesis of the polyether ionophore monensin. *Chembiochem : a European journal of chemical biology* **2006,** *7* (9), 1435-42.
16. Liu, T.; You, D.; Valenzano, C.; Sun, Y.; Li, J.; Yu, Q.; Zhou, X.; Cane, D. E.; Deng, Z., Identification of NanE as the thioesterase for polyether chain release in nanchangmycin biosynthesis. *Chemistry & biology* **2006,** *13* (9), 945-55.
17. Kalaitzis, J. A.; Cheng, Q.; Meluzzi, D.; Xiang, L.; Izumikawa, M.; Dorrestein, P. C.; Moore, B. S. J. B.; chemistry, m., Policing starter unit selection of the enterocin type II polyketide synthase by the type II thioesterase EncL. **2011,** *19* (22), 6633-6638.
18. Tang, Y.; Koppisch, A. T.; Khosla, C., The acyltransferase homologue from the initiation module of the R1128 polyketide synthase is an acyl-ACP thioesterase that edits acetyl primer units. *Biochemistry* **2004,** *43* (29), 9546-55.
19. Guntaka, N. S.; Healy, A. R.; Crawford, J. M.; Herzon, S. B.; Bruner, S. D. J. A. c. b., Structure and functional analysis of ClbQ, an unusual intermediate-releasing thioesterase from the colibactin biosynthetic pathway. **2017,** *12* (10), 2598-2608.
20. Heathcote, M. L.; Staunton, J.; Leadlay, P. F., Role of type II thioesterases: evidence for removal of short acyl chains produced by aberrant decarboxylation of chain extender units. *Chemistry & biology* **2001,** *8* (2), 207-20.
21. Pieper, R.; Ebert-Khosla, S.; Cane, D.; Khosla, C., Erythromycin biosynthesis: kinetic studies on a fully active modular polyketide synthase using natural and unnatural substrates. *Biochemistry* **1996,** *35* (7), 2054-60.
22. Jacobsen, J. R.; Cane, D. E.; Khosla, C., Spontaneous priming of a downstream module in 6-deoxyerythronolide B synthase leads to polyketide biosynthesis. *Biochemistry* **1998,** *37* (14), 4928-34.
23. Weissman, K. J.; Bycroft, M.; Staunton, J.; Leadlay, P. F., Origin of starter units for erythromycin biosynthesis. *Biochemistry* **1998,** *37* (31), 11012-7.
24. Pereda, A.; Summers, R. G.; Stassi, D. L.; Ruan, X.; Katz, L., The loading domain of the erythromycin polyketide synthase is not essential for erythromycin biosynthesis in Saccharopolyspora erythraea. *Microbiology (Reading, England)* **1998,** *144 (Pt 2)*, 543-53.

25. Schwarzer, D.; Mootz, H. D.; Linne, U.; Marahiel, M. A., Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases. *Proceedings of the National Academy of Sciences of the United States of America* **2002,** *99* (22), 14083-8.
26. Hu, Z.; Pfeifer, B. A.; Chao, E.; Murli, S.; Kealey, J.; Carney, J. R.; Ashley, G.; Khosla, C.; Hutchinson, C. R., A specific role of the Saccharopolyspora erythraea thioesterase II gene in the function of modular polyketide synthases. *Microbiology (Reading, England)* **2003,** *149* (Pt 8), 2213-2225.
27. Claxton, H. B.; Akey, D. L.; Silver, M. K.; Admiraal, S. J.; Smith, J. L., Structure and functional analysis of RifR, the type II thioesterase from the rifamycin biosynthetic pathway. *The Journal of biological chemistry* **2009,** *284* (8), 5021-9.
28. Miller, D. A.; Luo, L.; Hillson, N.; Keating, T. A.; Walsh, C. T., Yersiniabactin synthetase: a four-protein assembly line producing the nonribosomal peptide/polyketide hybrid siderophore of Yersinia pestis. *Chemistry & biology* **2002,** *9* (3), 333-44.
29. Chen, Y. L.; Zhao, J.; Liu, W.; Gao, J. F.; Tao, L. M.; Pan, H. X.; Tang, G. L., Identification of phoslactomycin biosynthetic gene clusters from Streptomyces platensis SAM-0654 and characterization of PnR1 and PnR2 as positive transcriptional regulators. *Gene* **2012,** *509* (2), 195-200.
30. Geyer, K.; Sundaram, S.; Susnik, P.; Koert, U.; Erb, T. J., Understanding Substrate Selectivity of Phoslactomycin Polyketide Synthase by Using Reconstituted in Vitro Systems. *Chembiochem : a European journal of chemical biology* **2020**.
31. Whicher, J. R.; Florova, G.; Sydor, P. K.; Singh, R.; Alhamadsheh, M.; Challis, G. L.; Reynolds, K. A.; Smith, J. L., Structure and function of the RedJ protein, a thioesterase from the prodiginine biosynthetic pathway in Streptomyces coelicolor. *The Journal of biological chemistry* **2011,** *286* (25), 22558-69.
32. Bruner, S. D.; Weber, T.; Kohli, R. M.; Schwarzer, D.; Marahiel, M. A.; Walsh, C. T.; Stubbs, M. T. J. S., Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE. **2002,** *10* (3), 301-310.
33. Cox, R. J.; Crosby, J.; Daltrop, O.; Glod, F.; Jarzabek, M. E.; Nicholson, T. P.; Reed, M.; Simpson, T. J.; Smith, L. H.; Soulas, F.; Szafranska, A. E.; Westcott, J., Streptomyces coelicolor phosphopantetheinyl transferase: a promiscuous activator of polyketide and fatty acid synthase acyl carrier proteins. *Journal of the Chemical Society, Perkin Transactions 1* **2002,** (14), 1644-1649.
34. Quadri, L. E.; Weinreb, P. H.; Lei, M.; Nakano, M. M.; Zuber, P.; Walsh, C. T., Characterization of Sfp, a Bacillus subtilis phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. *Biochemistry* **1998,** *37* (6), 1585-95.
35. Carreras, C. W.; Gehring, A. M.; Walsh, C. T.; Khosla, C., Utilization of enzymatically phosphopantetheinylated acyl carrier proteins and acetyl-acyl carrier proteins by the actinorhodin polyketide synthase. *Biochemistry* **1997,** *36* (39), 11757-61.
36. Bisang, C.; Long, P. F.; Cortes, J.; Westcott, J.; Crosby, J.; Matharu, A. L.; Cox, R. J.; Simpson, T. J.; Staunton, J.; Leadlay, P. F., A chain initiation factor common to both modular and aromatic polyketide synthases. *Nature* **1999,** *401* (6752), 502-5.
37. Long, P. F.; Wilkinson, C. J.; Bisang, C. P.; Cortes, J.; Dunster, N.; Oliynyk, M.; McCormick, E.; McArthur, H.; Mendez, C.; Salas, J. A.; Staunton, J.; Leadlay, P. F., Engineering specificity of starter unit selection by the erythromycin-producing polyketide synthase. *Molecular microbiology* **2002,** *43* (5), 1215-25.

38. Robbins, T.; Kapilivsky, J.; Cane, D. E.; Khosla, C., Roles of Conserved Active Site Residues in the Ketosynthase Domain of an Assembly Line Polyketide Synthase. *Biochemistry* **2016,** *55* (32), 4476-84.

Supplementary Information

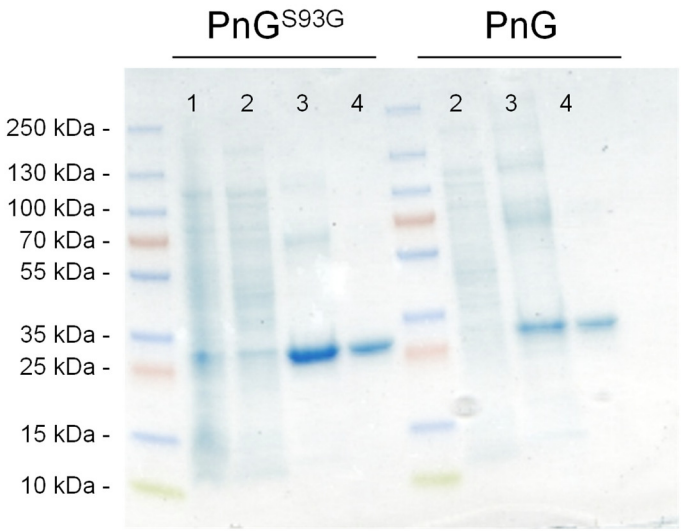


Figure S1: SDS-Page analysis of PnG and PnG S93G. Insoluble fraction (1), lysate (2), Ni-NTA agarose elution (3) and the purified protein after size exclusion chromatography (4). Expected size of PnG is 30.2 kDa.

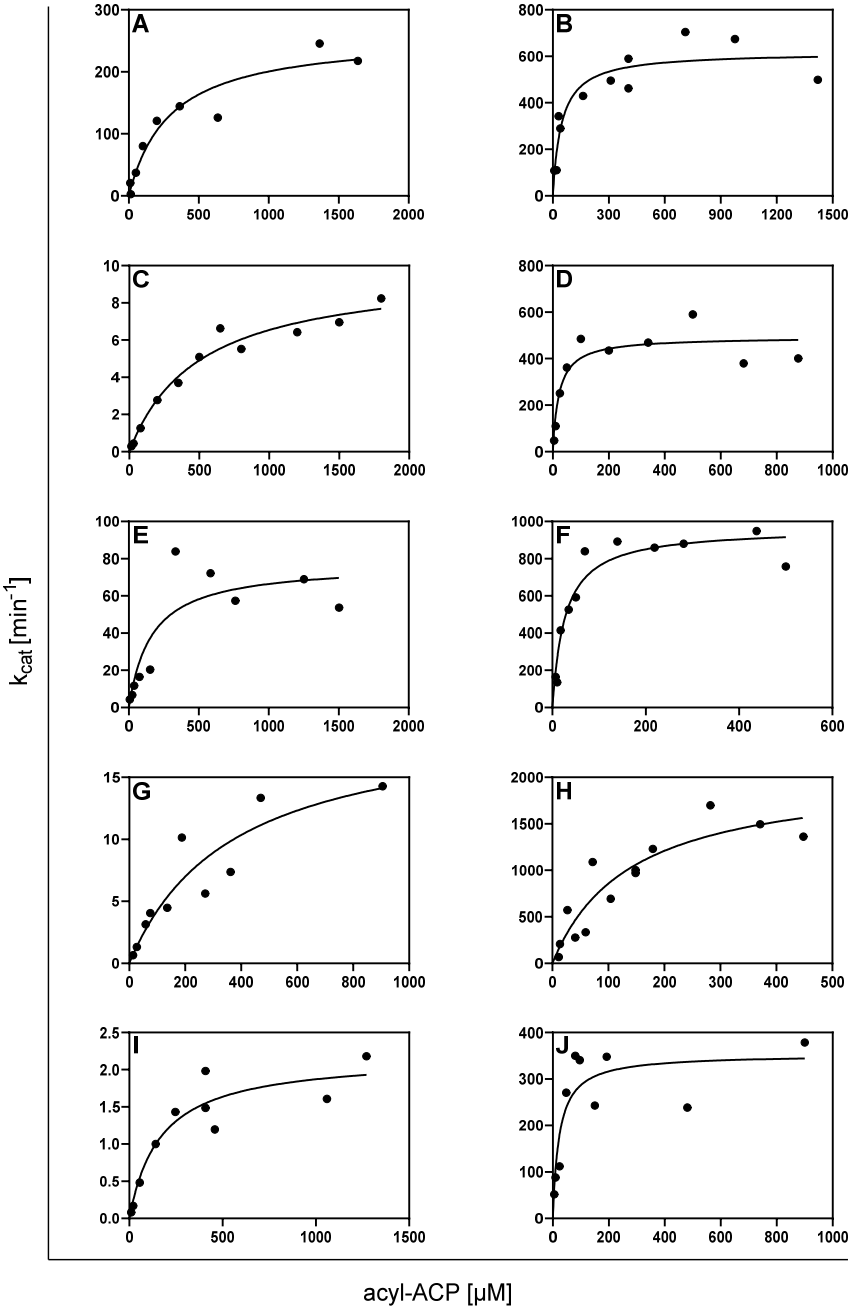


Figure S2: Kinetic measurements of PnG hydrolytic activity towards ACP bound acyl-residues fitted to Michaelis Menten plot (GraphPad Prism 8.0.0) **A)** malonyl-ACP_{LD}, **B)** acetyl-ACP_{LD}, **C)** ethylmalonyl-ACP_{LD} **D)** butyl-ACP_{LD}, **E)** malonyl-ACP2, **F)** acetyl-ACP2, **G)** methylmalonyl-ACP2, **H)** propionyl-ACP2, **I)** ethylmalonyl-ACP2, **J)** butyl-ACP2.

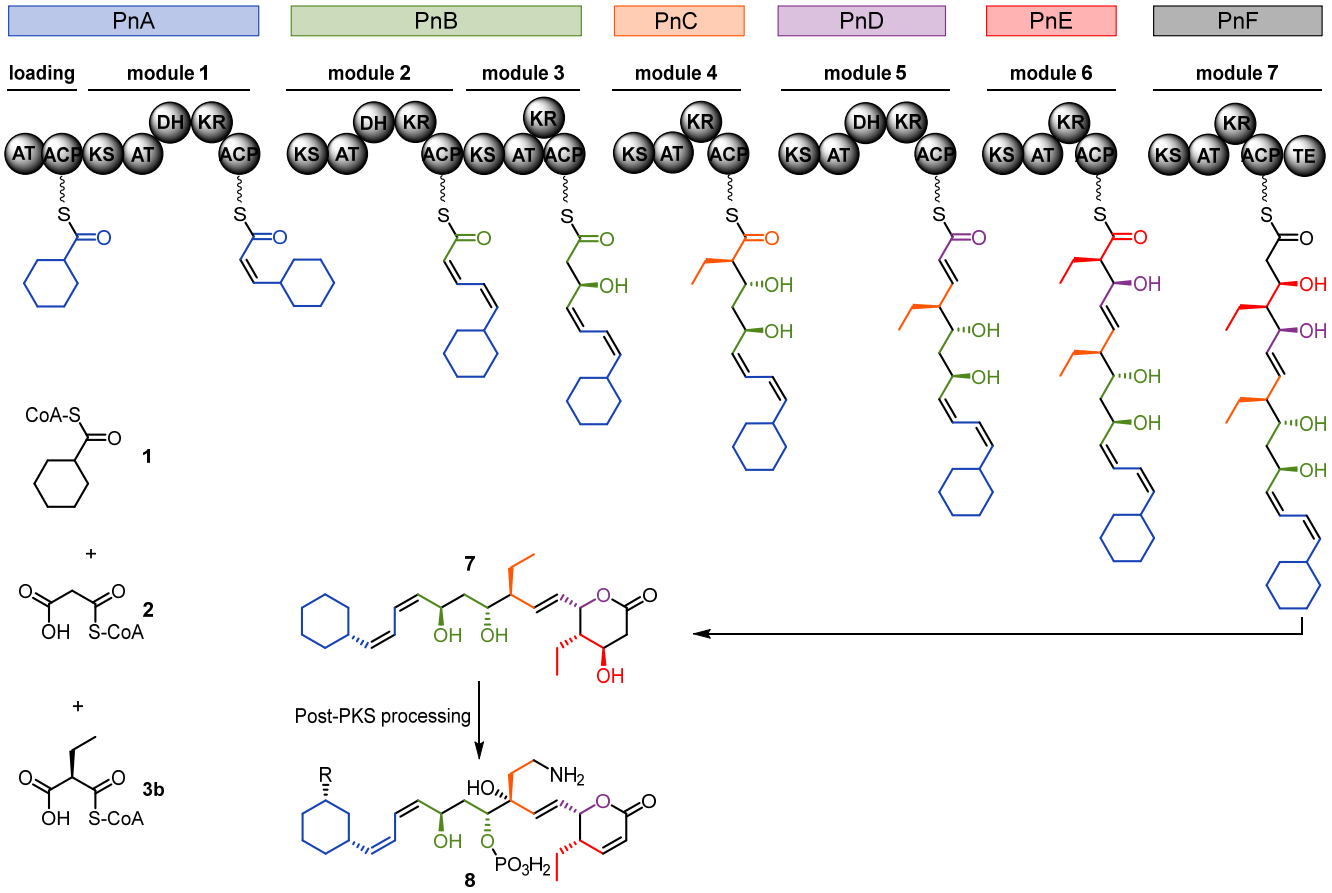


Figure S3: Phoslactomycin (Pn) PKS. ACP=acyl carrier protein; AT=acyltransferase; DH=dehydratase; KR=ketoreductase; TE=thioesterase; **1**=cyclohexanecarboxyl-CoA; **2**=malonyl-CoA; **3**=(2*S*)-ethylmalonyl-CoA; **7**=phoslactomycin polyketide backbone, **8**=bioactive phoslactomycin derivatives; **R**=isobutyloxy; isovaleryloxy; 4-methylcaleryloxy; cyclohexylcarbonyloxy; 4-methylheptanoyloxy. [1, 2].

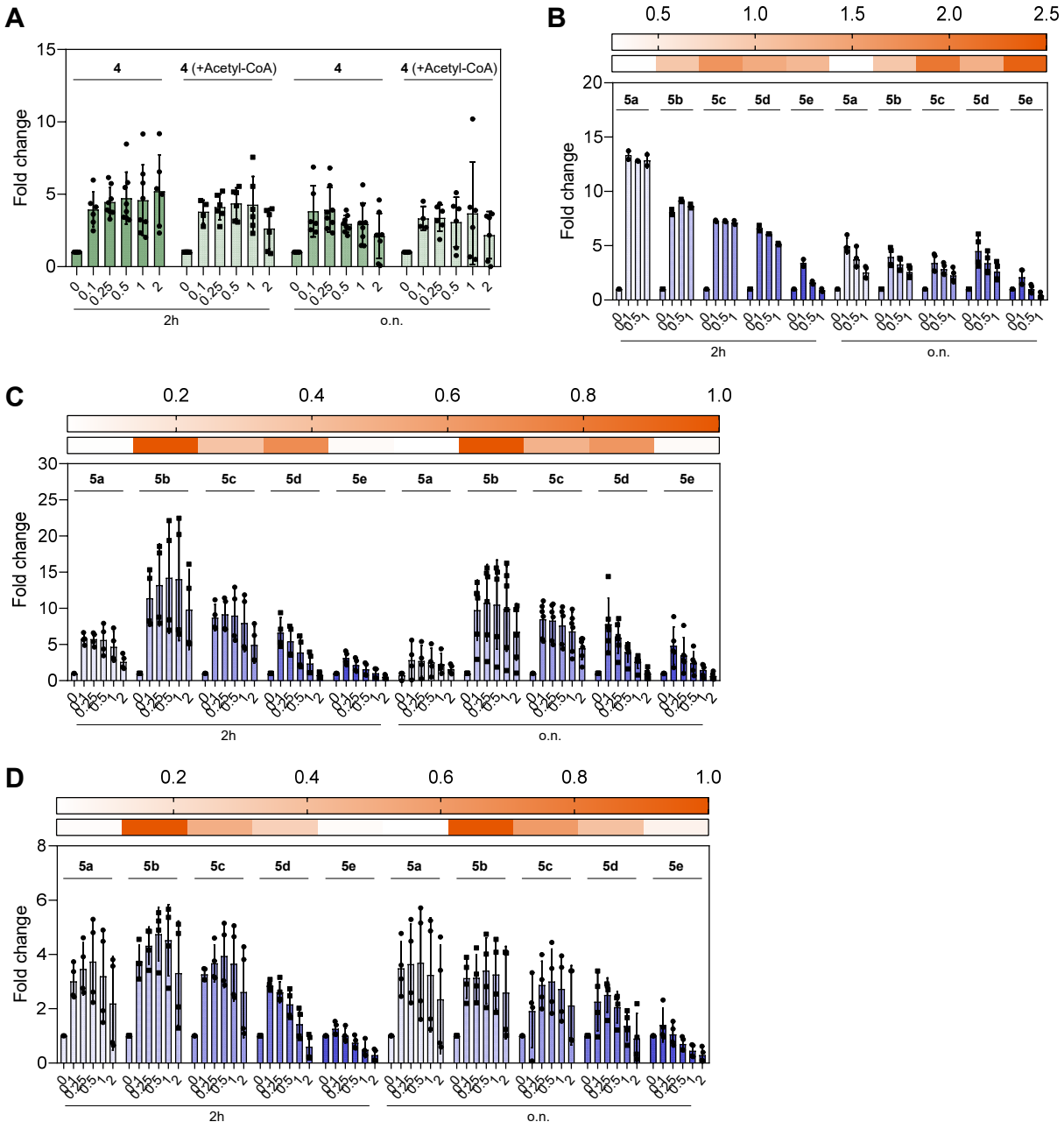


Figure S4: Effect of increasing PnG concentration on polyketide production after 2 h and overnight. All sample series were normalized to the control sample without PnG. Concentrations of PnG, indicated on the x-axis ranged between 0.1 and 2 molar equivalents of ACPs present in the assay. In **B**, **C** and **D** the heat map indicates the initial product distribution, set relative to the natural product **5b**. Residues indicated are: **a**=methyl-, **b**=ethyl-, **c**=butyl-, **d**=3-methlbutyl-, **e**=hexyl-residues. (**A**) Tetraketide production is increased up to five-fold. No effect of acetyl-CoA addition can be observed. (**B**) Pentaketide production in reaction mixtures containing malonyl-CoA and one *α*-substituted malonyl-CoA derivative as extender units. Increase in product formation can be seen for all pentaketide derivatives **5a-e**. The strong increase of **5a**

can be explained by the very little initial amounts close to the detection limit. The overall lower fold-changes in the overnight samples could be explained by the consecutive biosynthesis of pentaketides in the samples not containing PnG, while samples containing PnG ran out of resources. (**C**) Pentaketide production run under competitive reaction conditions, containing malonyl-CoA and all five *α*-substituted malonyl-CoA extender units. Initial product amount of **5b** is highest, as well as the fold-change when PnG is added. (**D**) Pentaketide production with pre-treated Pn PKS enzymes. Treatment with PnG ensures removal of all potentially aberrant acyl-residues. PnG is removed by size exclusion chromatography from the Pn PKS enzymes. Data points are from multiple biological (**A**, **B** o.n, **C**, **D**) and technical (**A**, **B**, **C**, **D**) replicates.

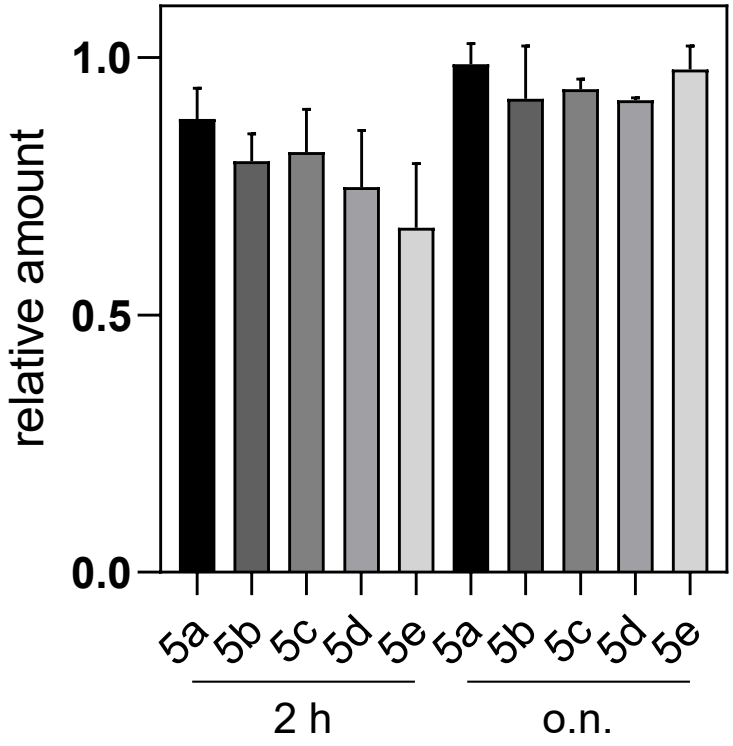
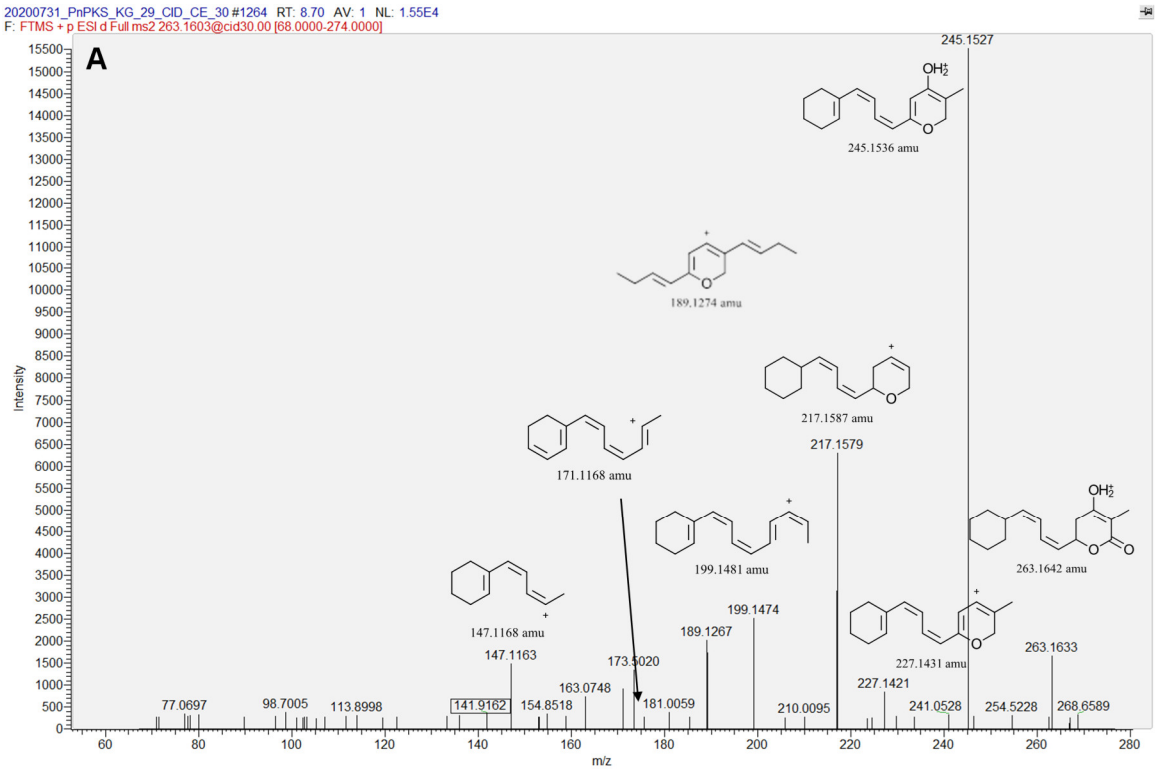
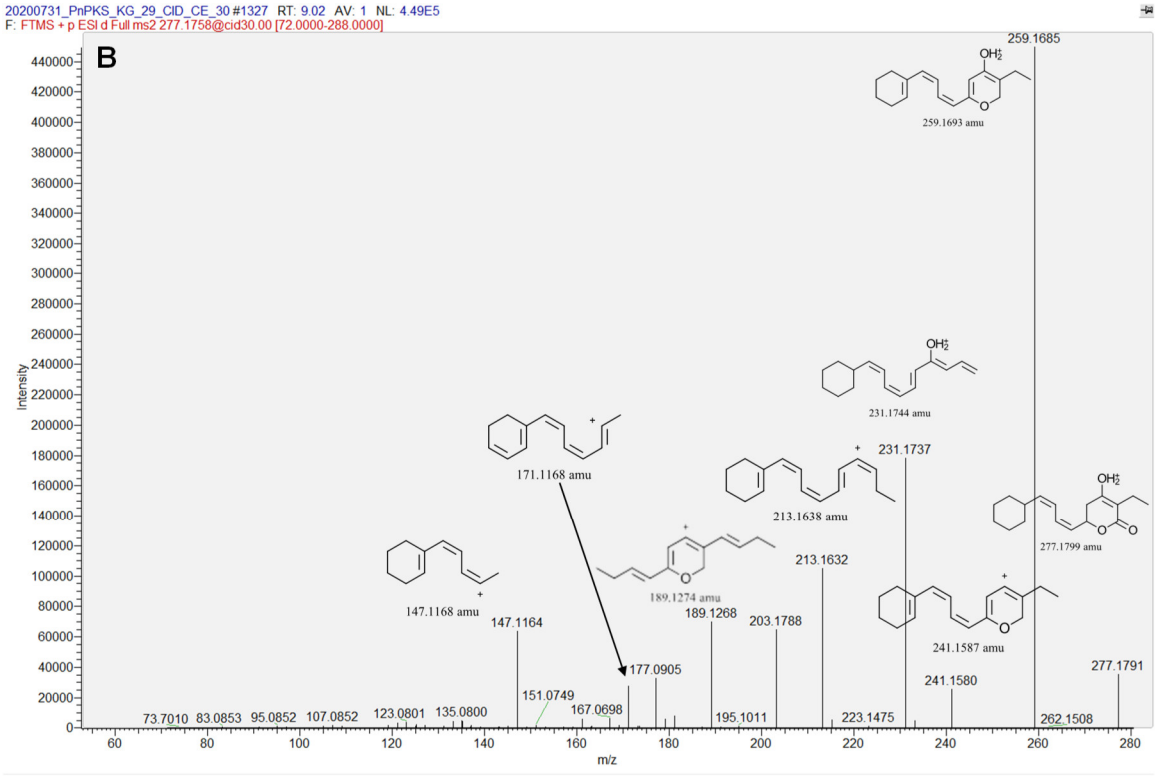
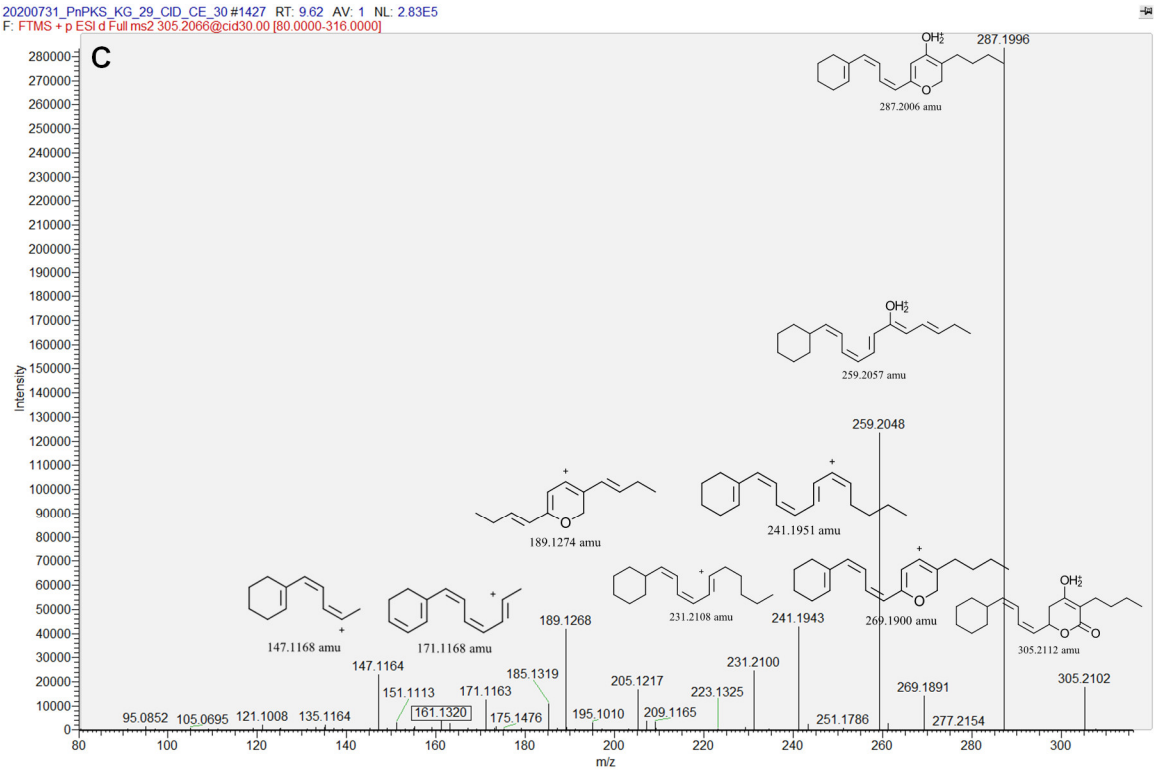
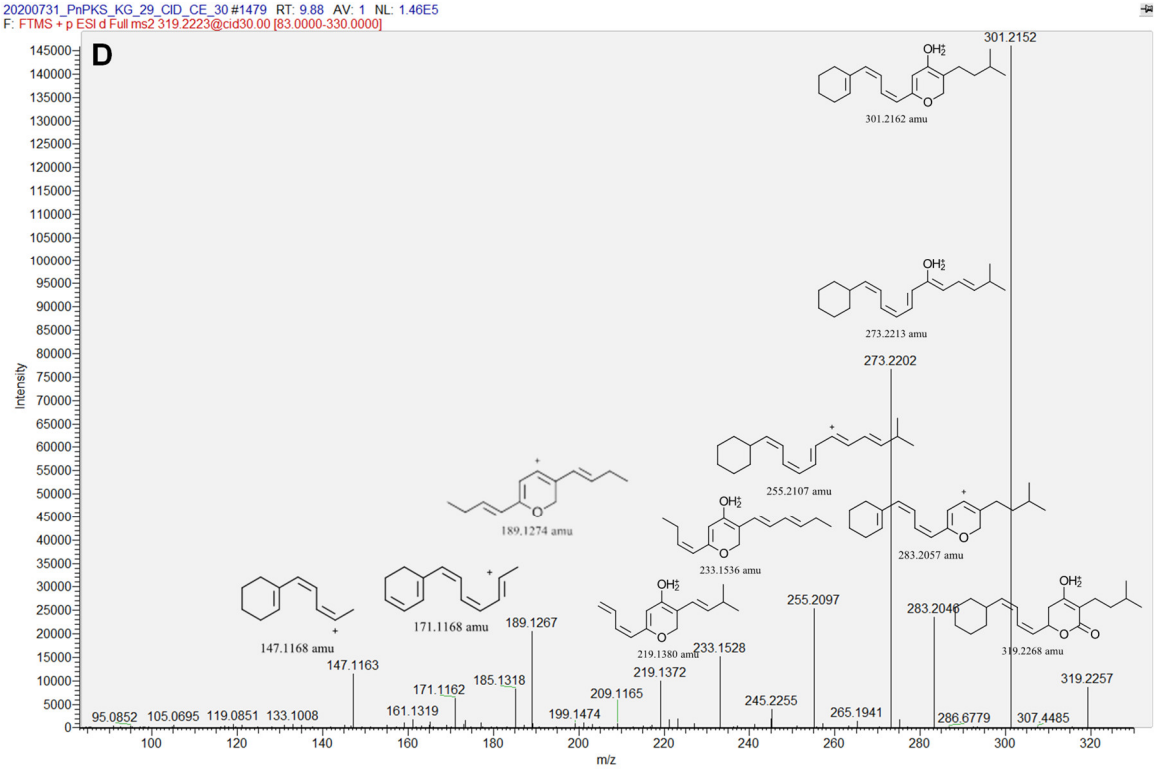


Figure S5: PnG catalytic knockout control PnG S93G. The active site serine is mutated to a glycine, rendering it unable to covalently bind the substrate. This control was performed to assure that all effects, observed in the assays are due to PnG and not caused by protein contaminations. Shown here is the relative product amount of competitive pentaketide assays containing PnG S93G compared with assays containing no PnG. Samples containing PnG S93G, taken after 2 h and o.n., contain the same product amounts as samples without the type II thioesterase PnG. Residues indicated are: **a**=methyl-, **b**=ethyl-, **c**=butyl-, **d**=3-methlbutyl-, **e**=hexyl-residues.


20200731_PnPKS_KG_29_CID_CE_30 #1264 RT: 8.70 AV: 1 NL: 1.55E4
F: FTMS + p ESI d Full ms2 263.1603@cid30.00 [68.0000-274.0000]
A
Intensity
m/z
245.1527
245.1536 amu
189.1274 amu
217.1587 amu
217.1579
171.1168 amu
199.1481 amu
263.1642 amu
147.1168 amu
199.1474
189.1267
263.1633
227.1431 amu
147.1163
173.5020
227.1421
163.0748
77.0697
98.7005
113.8998
141.9162
154.8518
181.0059
210.0095
241.0528
254.5228
268.6589


20200731_PnPKS_KG_29_CID_CE_30 #1327 RT: 9.02 AV: 1 NL: 4.49E5
F: FTMS + p ESI d Full ms2 277.1758@cid30.00 [72.0000-288.0000]
B
Intensity
m/z
259.1685
259.1693 amu
231.1744 amu
231.1737
171.1168 amu
213.1638 amu
277.1799 amu
213.1632
189.1274 amu
147.1168 amu
189.1268
203.1788
147.1164
241.1587 amu
277.1791
177.0905
241.1580
151.0749
167.0698
73.7010
83.0853
95.0852
107.0852
123.0801
135.0800
195.1011
223.1475
262.1508


20200731_PnPKS_KG_29_CID_CE_30 #1427 RT: 9.62 AV: 1 NL: 2.83E5
F: FTMS + p ESI d Full ms2 305.2066@cid30.00 [80.0000-316.0000]
C
Intensity
m/z
287.1996
287.2006 amu
259.2057 amu
259.2048
189.1274 amu
241.1951 amu
147.1168 amu
171.1168 amu
189.1268
231.2108 amu
241.1943
269.1900 amu
305.2112 amu
147.1164
151.1113
171.1163
185.1319
205.1217
231.2100
223.1325
269.1891
305.2102
95.0852
105.0695
121.1008
135.1164
161.1320
175.1476
195.1010
209.1165
251.1786
277.2154


20200731_PnPKS_KG_29_CID_CE_30 #1479 RT: 9.88 AV: 1 NL: 1.46E5
F: FTMS + p ESI d Full ms2 319.2223@cid30.00 [83.0000-330.0000]
D
Intensity
m/z
301.2152
301.2162 amu
273.2213 amu
273.2202
255.2107 amu
189.1274 amu
233.1536 amu
283.2057 amu
147.1168 amu
171.1168 amu
189.1267
219.1380 amu
233.1528
255.2097
283.2046
319.2268 amu
147.1163
171.1162
185.1318
219.1372
209.1165
245.2255
319.2257
95.0852
105.0695
119.0851
133.1008
161.1319
199.1474
265.1941
286.6779
307.4485

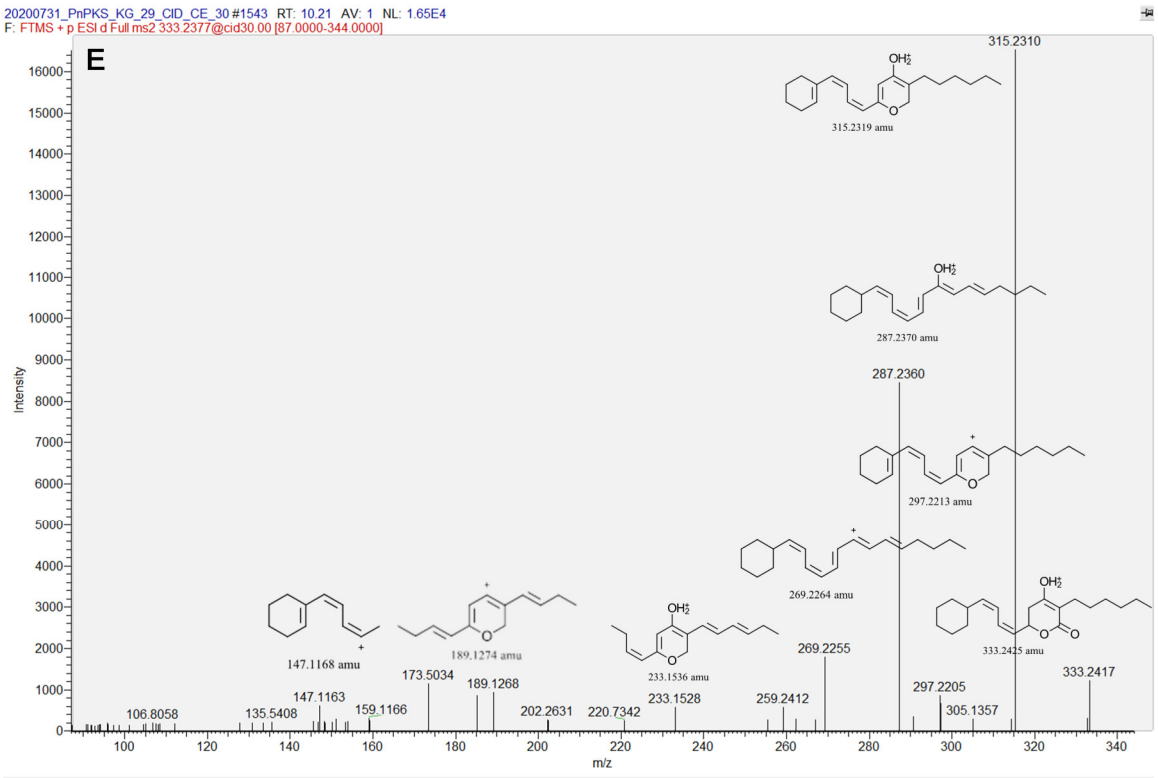


Figure S6: Tandem mass spectrometric analysis of native and non-native pentaketides

Putative structures for the different polyketides were generated using knowledge on the biosynthetic logic of the phoslactomycin PKS system. The fragment masses were obtained by inspecting the data files using Xcalibur Qual Browser software. The experimental fragments were used to verify the proposed structure of the molecules by drawing fragment structures to explain the observed fragment mass. The structures of the fragments were drawn using ChemDoodle software. All precursor ion and fragment ion masses differed by less than 5 ppm compared to the theoretical mass. In addition, *in-silico* fragmentation was performed using MetFrag online. However, the top scoring candidates had molecular formulas that are unnatural (eg. Contains Boron or Fluorine, or Halogen) and are not expected to be formed by the Pn PKS *in vitro*. **A)** Fragmentation pattern of **5a** (methylmalonyl-CoA incorporation), **B)** Fragmentation pattern of **5b** (native product, ethylmalonyl-CoA incorporation, **C)** Fragmentation pattern of **5c** (butylmalonyl-CoA incorporation), **D)** Fragmentation pattern of **5d** (3-methylbutylmalonyl-CoA incorporation) **E)** Fragmentation pattern of **5e** (hexylmalonyl-CoA incorporation)

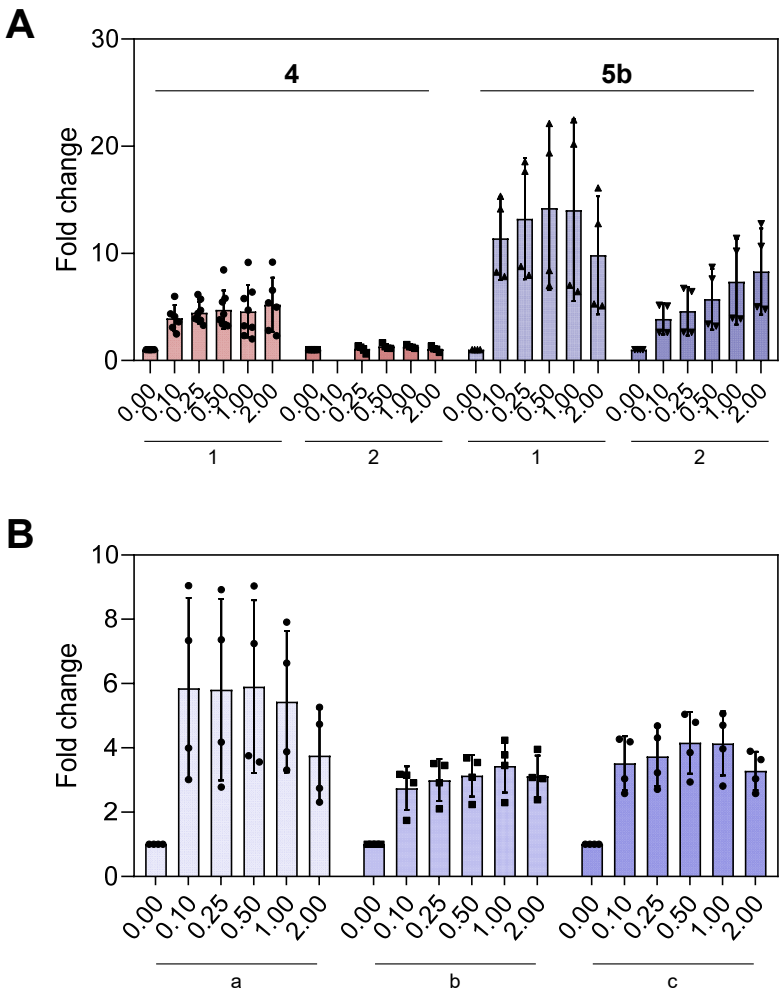


Figure S7: **Diketide SNAC thioester impacts PnGs activity**. The x-axis in **A** and **B** indicates the concentration of PnG, equimolar towards the ACPs present, samples taken after 2 h. (**A**) Pn polyketide production can be initiated by PnA_{V4} (panel 1). Alternatively, PnA_{V4} can be bypassed when the respective diketide SNAC analog, serving as a substrate for PnB is added (panel 2). When initiated with PnA_{V4} tetraketide production increased five-fold, surprisingly when bypassing PnA_{V4} and supplying the SNAC ester, no effect of PnG was detected. When producing the pentaketide **5b**, a 14-fold increase in production upon PnG addition is observed, however when bypassing PnA_{V4} **5b** production increases maximal seven-fold. Either PnG has a differential effect on PnA_{V4}, PnB and PnC-TE_{DEBS} or the SNAC-ester has an inhibition-like effect on PnG. (**B**) To test if the diketide-SNAC analog is responsible for the decrease of PnGs effect on polyketide production, we ran **5b** production assays with PnA_{V4}, PnB, PnC-TE_{DEBS} (a), PnB, PnC-TE_{DEBS} (b) and PnA_{V4}, PnB, PnC-TE_{DEBS} plus the diketide-SNAC ester (c). A six-fold increase was observed in a, however in b and c an increase of only approximately four-fold was measured. This leads to the assumption that the diketide SNAC analog impacts PnG activity. SNAC thioesters also serve as substrates for thioesterases, however with a slower turnover. PnG most likely binds the SNAC ester as substrate leading to a less strong effect on Pn PKS.

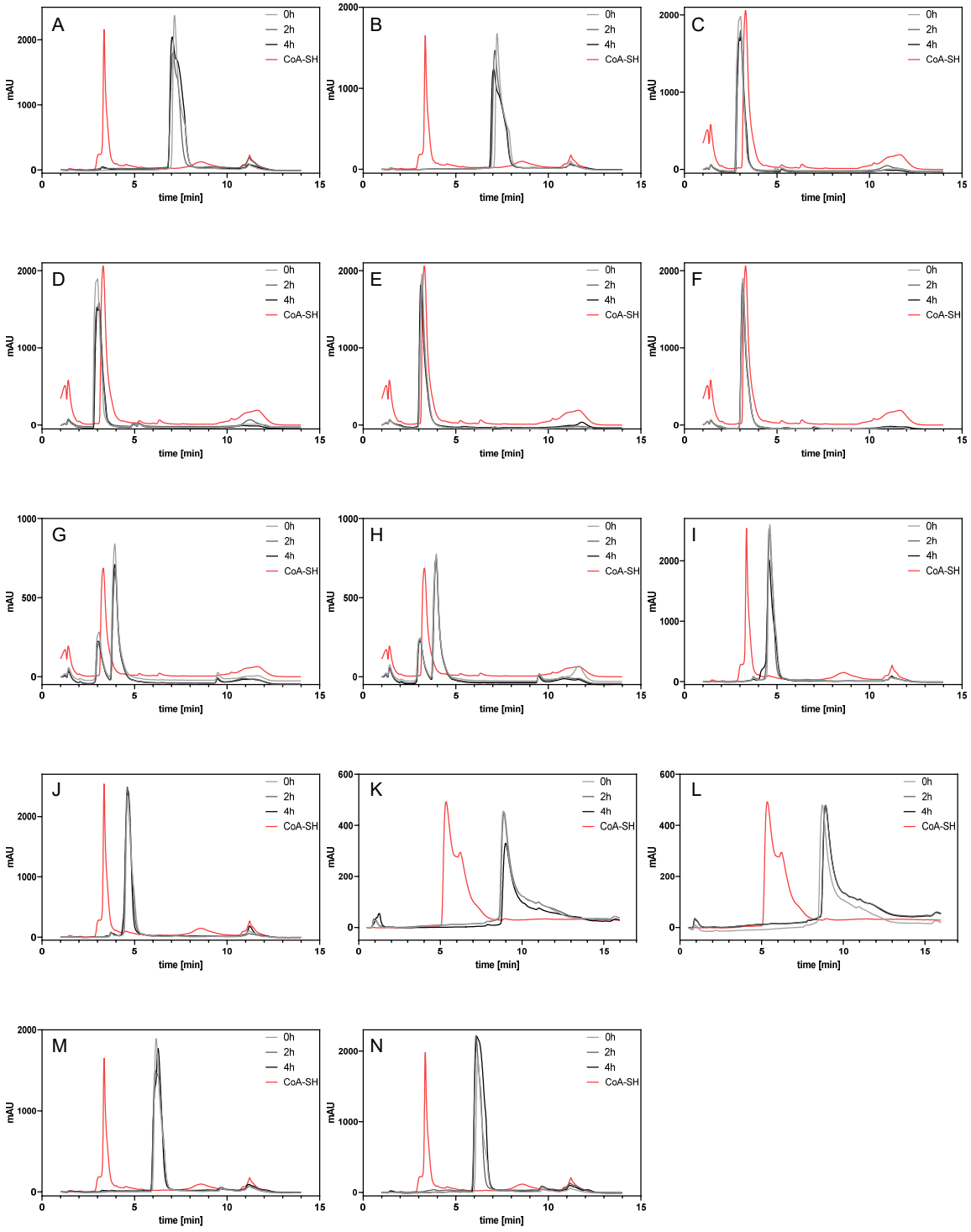


Figure S8: PnG does not hydrolyze CoA-thioesters. Incubation of PnG with the CoA-thioester substrates present in the polyketide production assays. Samples from assays containing PnG and without PnG as negative control were taken after 0 h, 2 h and 4 h. Neither buildup of free CoA-SH, nor depletion of the acyl-CoA thioester can be observed. **A)** Cyclohexanecarboxyl-CoA with PnG, **B)** and without PnG. **C)** Malonyl-

CoA with PnG, **D)** and without PnG. **E)** Methylmalonyl-CoA with PnG, **F)** and without PnG. **G)** Ethylmalonyl-CoA with PnG, **H)** and without PnG. **I)** Butylmalonyl-CoA with PnG, **J)** and without PnG. **K)** 3-methylbutylmalonyl-CoA with PnG, **L)** and without PnG. **M)** Hexylmalonyl-CoA with PnG, **N)** and without PnG.

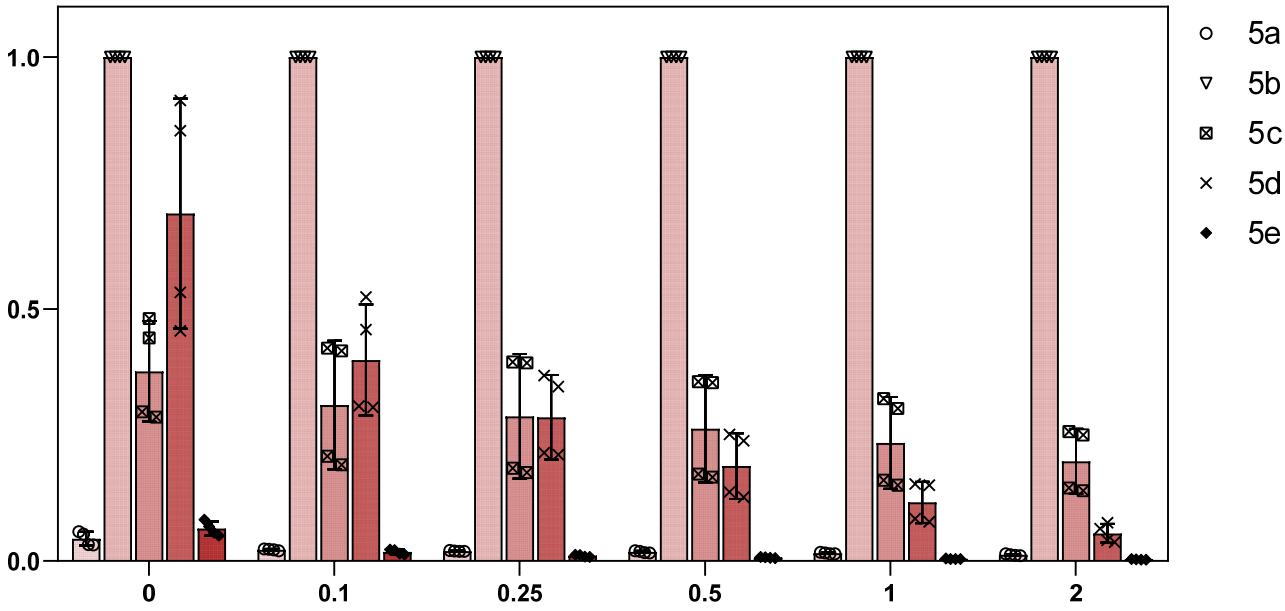


Figure S9: Effect of PnG on the product profile in competitive pentaketide assays. All samples are normalized to the **5b** ion count in the same sample. When comparing the products in competitive assays it shows that all alternative products relatively decrease upon addition of PnG (while the absolute amount of product increases for all – compare to Figure S4). This shaping in product profile towards the natural product **5b** could be caused a longer exposure of non-native acyl-residues due to slower processing. The longer exposure results in a higher probability of PnG to hydrolyze the residue, which increases with higher PnG concentration. Shown are the values for two biological replicate with each two technical replicates.

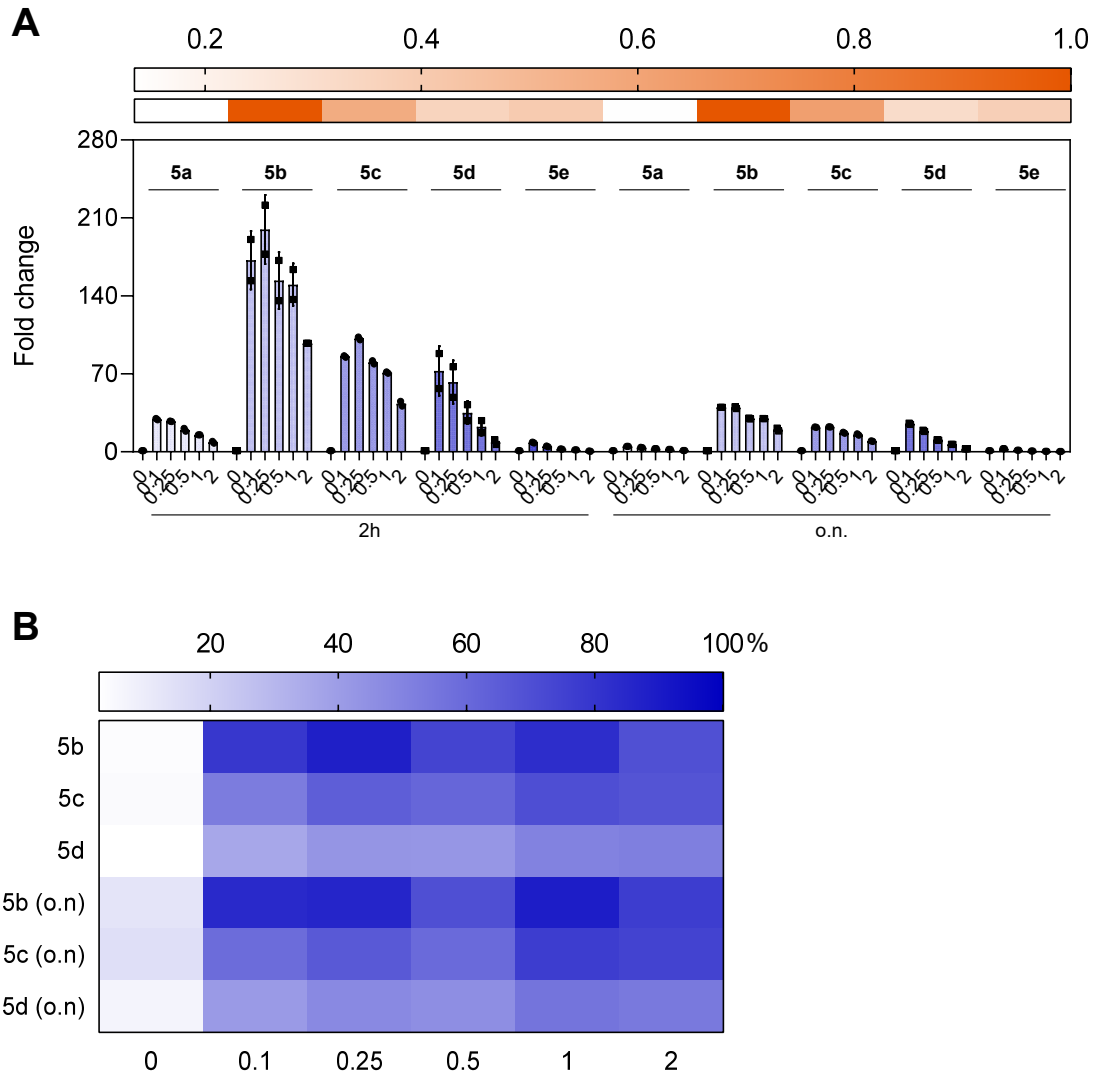


Figure S10: Effect of PnG on the pentaketide production and release without terminal thioesterase. Pn PKS assembly line is missing the terminal thioesterase (PnA_{V4}-PnB-PnC). Residues indicated are: **a**=methyl-, **b**=ethyl-, **c**=butyl-, **d**=3-methlbutyl-, **e**=hexyl-residues. Concentrations of PnG ranged between 0.1 and 2 molar equivalents of ACPs present in the assay. Pentaketide production was run under competitive reaction conditions, all mixtures contained malonyl-CoA and all *α*-substituted extender units. **A)** Pentaketide production after 2 h and overnight (o.n.) incubation of the assay. All sample series were normalized to the control sample without PnG. The heat map displays the product distribution of non-native pentaketides (**5a**, **5c-d**) compared to the native pentaketide **5b** in absence of PnG. A strong increase of product formation after 2 h can be observed for all pentaketides, particularly **5b-5c** (200- to 100-fold increase respectively). The effect of PnG on product formation becomes less prominent in the overnight samples. This is most likely due to slow but consistent production of pentaketides in the assays not containing PnG. **B)** Heat map of production comparison of assays terminating without (PnA_{V4} + PnB + PnC)

and with a terminal thioesterase (PnA_{V4} + PnB + PnC-TE_{DEBS}) to identify if PnG can functionally reconstitute the terminal TE, samples after 2 h and overnight (o.n.). In absence of PnG product formation reaches only 5% of the level when a terminal TE is present. This supports the importance of the fusion of TE_{DEBS} to PnC for pentaketide release. Upon addition of PnG final product amount in ion counts reaches up to approx. 90% (**5b**), 80% (**5c**) and 55% (**5d**) of that, if TE_{DEBS} was present. PnG is able to hydrolyze the pentaketide product bound to PnC.

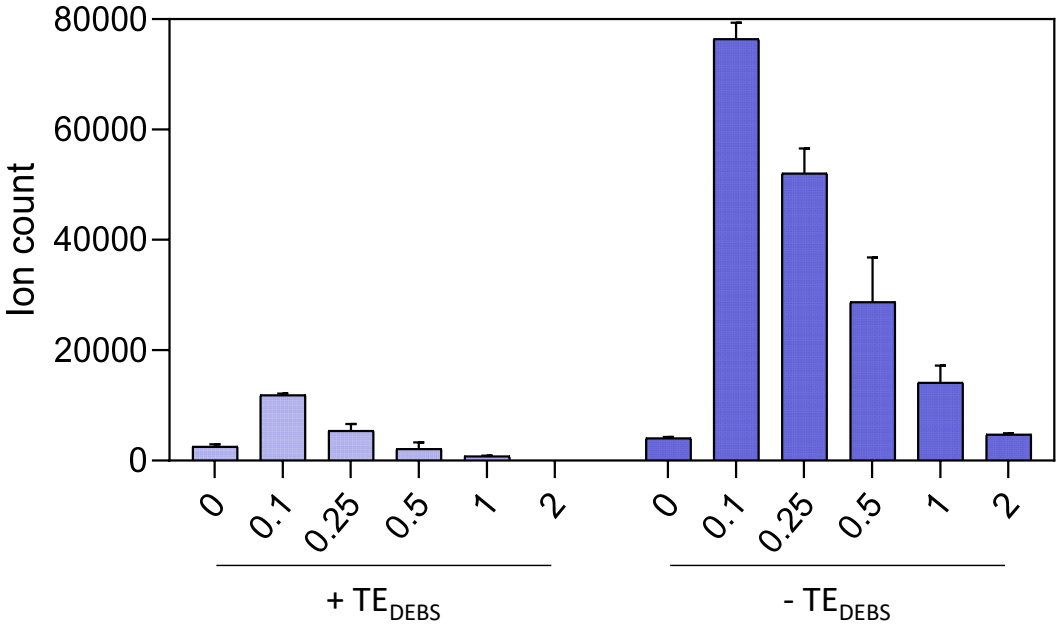


Figure S11: Release of hexaketides by PnG. The x-axis shows the concentration of PnG, equimolar to ACPs present in the assay. On the left side, PnG has been added to assays terminating with TE_{DEBS} (PnA_{V4}, PnB, PnC, PnD-TE_{DEBS}), on the right side without TE_{DEBS} (PnA_{V4}, PnB, PnC, PnD). A five-fold increase in product formation can be observed upon addition of PnG to assays terminating with TE_{DEBS}, while a 20-fold increase can be observed in absence of TE_{DEBS}. In the control samples without PnG, assays terminating without TE_{DEBS} contain two-fold higher product amounts then when a terminal TE is present. This shows a negative influence of the terminal TE from DEBS on Pn hexaketide production.

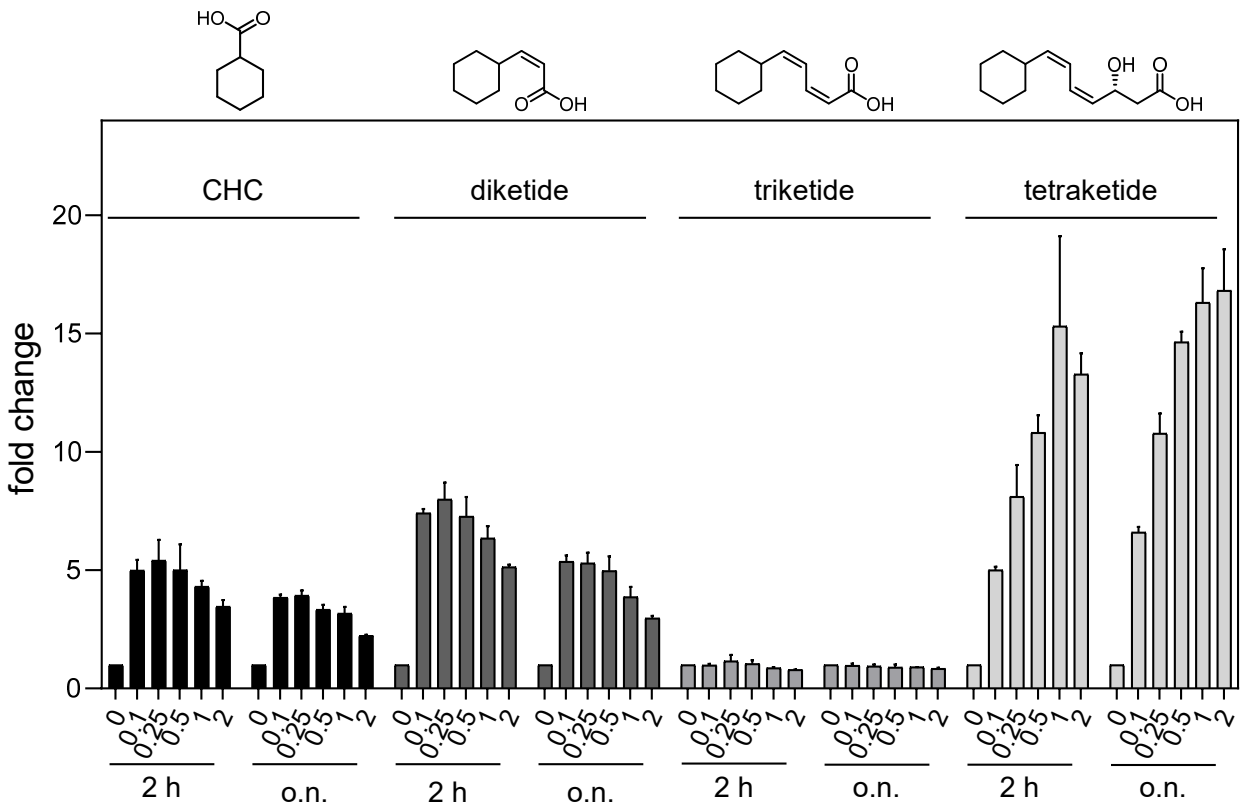


Figure S12: Release of intermediates from Pn PKS pentaketide system. Initiation with PnA_{V4} and termination with PnC-TE_{DEBS}. Shown are the peak areas of samples with increasing PnG concentration relative to the control without PnG. Concentrations of PnG ranged between 0.1 and 2 molar equivalents of ACPs present in the assay. CHC=cyclohexanecarboxylic acid ($C_7H_{12}O_2$; calculated: *m/z* [H^+]: 129.091; *m/z* [Na^+]: 151.073; found: *m/z* [Na^+]: 151.0758; retention time 8.5 min), diketide ($C_9H_{14}O_2$; calculated: *m/z* [H^+]: 155.1067; *m/z* [Na^+]: 177.0886; found: *m/z* [Na^+]: 177.0926; retention time 8.5 min), triketide ($C_{11}H_{16}O_2$; calculated: *m/z* [H^+]: 181.1223; *m/z* [Na^+]: 203.1043; found: *m/z* [Na^+]: 203.1088; retention time 9 min), tetraketide ($C_{13}H_{20}O_3$; calculated: *m/z* [H^+]: 225.1485 ; *m/z* [Na^+]: 247.1305; dehydrated $C_{13}H_{18}O_2$; *m/z* [H^+]: 207.138; *m/z* [Na^+]: 229.1199; found *m/z* [H^+] 207.1397; retention time 7 min). No difference can be observed between the 2 h and overnight (o.n.) samples. Release of triketide cannot be observed. The tri- and tetraketide are produced by the bi-modular protein PnB. The triketide is internally translocated to the downstream module, protecting it from PnG.

Table S1: List of polyketide masses (*m/z*) from the *in vitro* production assays. Both, *m/z* calculated and detected are given together with the retention time. **4** corresponds to the tetraketide, **5** to the pentaketide. **a** indicates methyl-, **b** ethyl-, **c** butyl-, **d** 3-methyl-butyl and **e** hexyl-residue.

Compound	Retention time [min]	$[H^+]$ *m/z* calculated	$[H^+]$ *m/z* detectedd	Δ *m/z*	$[Na^+]$ *m/z* calculated	$[Na^+]$ *m/z* detectedd	Δ *m/z*
4	7.1	207.1380	207.1403	0.0023	229.1199	-	
5a	8.1	263.1642	263.1648	0.0006	285.1461	285.1316	0.0145
	8.1	245.1536	245.1546	0.0010	267.1355	-	
5b	8.5	277.1798	277.1818	0.0020	299.1618	299.1635	0.0017
	8.5	259.1693	259.1710	0.0017	281.1512	-	
5c	9.2	305.2111	305.2130	0.0019	327.1931	327.1948	0.0017
	9.2	287.2006	287.2025	0.0019	309.1825	-	
5d	9.5	319.2268	319.2287	0.0019	341.2087	341.2105	0.0018
	9.5	301.2162	301.2180	0.0018	323.1982	-	
5e	9.8	333.2424	333.2439	0.0015	355.2244	355.2313	0.0069
	9.8	315.2319	315.2330	0.0011	337.2138	-	

Experimental section

Synthesis of substrates. Acyl-CoA synthesis was done like previously described. CHC-CoA was synthesized by chemical CDI coupling of the free acid, analogous as previously explained [3]. The diketide-SNAC was synthesized as previously reported [4]

Expression plasmid cloning. Preparation of the constructs for expression of PnA_{V4}, PnB, PnC and $PnC\text{-}TE_{DEBS}$ was reported before [4]. PnG coding sequence (optimized for expression in *E. coli*):

ATGGTGACCGAAAGCGATCGCTGGATTCGCAGCTATCTGCCGGGCCCGGCGGAA
GCGCTGCCGGTGGTGCATTTTCCGCATGCGGGCGGCAGCGCGAGCTATTATCGC
CCGTTTTGCGCGGCGCTGAGCGATCGCTTTAACGCGCTGGCGCTGCAGTATCCGG
GCCGCCAGGATCGCCGCGATGAACCGTGCGTGACCGATCTGCATGTGCTGGCGG
ATCTGCTGTTTGATCGCCTGCGCCAGTGCGCGGATCGCCCGATGGCGTTTTTTGG
CCATAGCATGGGCGCGCTGCTGGCGTTTGAAGTGACCCGCCGCTTTGAACGCGAA
CTGAACACCAGCCCGGTGGCGCTGTTTCTGAGCGGCCGCCGCGCGCCGAGCCGC
CATCGCGATGAAAACGTGGATCTGAGCAGCAACGAAAGCCTGCTGGCGGAAATTC
GCGAACTGAGCGGCACCGATCCGCGCCTGCTGGGCGATGATGAAATGCTGGAAA
TGATTATGGAACCGCTGCGCGCGGATTATCAGGCGCTGGGCGCGTATCATTTTGC

GCCGGAACCGCCGGTGCGCTGCCCGGTGACCGTGCTGACCGGCGCGGATGATC
CGCGCACCAGCCAGGATGAAGCGGCGGCGTGGCAGGAACATACCACCGGCGCG
TTTGATCTGCGCGTGTTTCCGGGCGGCCATTTTTTTATTAGCGAAAACGTGGCGGA
TGTGACCGGCTTTGTGGCGGAACGCCTGAGCGCGGTGCCGCTGGCGGGCTAA

The PnG coding sequence was inserted between the NdeI and HindIII multiple cloning sites of pET28b(+). The ordered sequence contained 5'- and 3'- overhangs of 8bp length to enable restriction by NdeI and HindII. pET28b(+) was digested with the same overhangs and the PnG and linearized pet28b(+) was purified from 0.8% agarose gel and ligated by T4 ligase mediated reaction. Single colonies were picked, grown, the plasmid isolated and sequenced.

Protein production and Nickel NTA-purification. Expression constructs encoding for the Pn PKS were transformed in *E. coli* BAP1 competent cells, all other proteins used in this study, including standalone ACP domains, were transformed in *E. coli* BL21 (DE3) competent cells and supplemented with the respective antibiotics. For PnG 1 L Terrific Broth medium was inoculated from plate and grown at 37 °C until OD_{600} reached 1.8. The culture was placed shaking at 23 °C and expression was induced with 100 µM isopropyl-β-D-thiogalactopyranoside. After 4 h of incubation, the culture was harvested and used for protein purification or stored at -20 °C. For all other expressions a 10 mL overnight culture was grown and used to inoculate 1 L Terrific Broth medium. The cultures were grown at 37 °C until an OD_{600} of 1 was reached, then placed shaking at 18 °C (Pn PKS proteins) or 20 °C overnight. The cultures were harvested and used for protein purification or stored at -20 °C. Cell pellets were resuspended in Buffer A (500 mM NaCl, 50 mM NaH_2PO_4, 10% v/v glycerol, pH 7.5) and lysed by sonication, centrifuged at 42.000 g, 4 °C for 45 min and the supernatant was mixed with 3.5 mL Protino® Ni-NTA Agarose purchased from Macherey Nagel and incubated for 2 h on ice with slow shaking. The beads solution was transferred into Protino® Columns 14 mL, washed with 50 mL Buffer A. Following the beads were washed with 20 mL washing buffer (25 mM Imidazole, 500 mM NaCl, 50 mM NaH_2PO_4, 10% v/v glycerol, pH 7.5). Proteins were eluted with 8 mL Buffer B (500 mM Imidazole, 500 mM NaCl, 50 mM NaH_2PO_4, 10% v/v glycerol,

pH 7.5). PnG and all standalone ACP proteins were further purified by size exclusion using HiLoad 16/600 Superdex 200 pg column (300 mM NaCl, 25 mM NaH_2PO_4, pH 7.5). Pn PKS proteins were subjected to anion exchange. Protein concentration was determined by UV-vis measurements at 280 nm, in the case of ACPs concentration was determined by Bradford assay.

Anion exchange. The eluate from the Nickel beads purification was diluted to 120 mL with Anion A (50 mM NaH2PO4, pH 7.5) and loaded onto a 5 mL HiTrap Q HP anion exchange 5 mL chromatography column, purchased from GE Healthcare Life Sciences, with a flow of 3 mL/min. A gradient to 100% Anion B (50 mM NaH_2PO_4, 500 mM NaCl pH 7.5) with a flow of 4 mL/min in 30 min was run and protein containing fractions were collected and concentrated using Amicon® Ultra Centrifugal Filters, purchased from Merck Millipore. Proteins were used immediately for assays or stored with 30% v/v glycerol in -80 °C after shock freezing in liquid nitrogen.

Pn PKS and PnG *in vitro* assays. The production of phoslactomycin polyketide derivatives was initiated using PnA_{V4} and cyclohexanecarboxyl-CoA. The assay was run at a volume of 50 µL and contained the 5 µM Pn PKS proteins, 3.5 µM Npt, 1 mM cyclohexanecarboxyl-CoA, 1 mM malonyl-CoA and 0.5 mM *α*-substituted malonyl-CoA derivatives (in single extender unit assays all extender units were used at 1 mM concentration), 5 mM NADPH, 0.1 mM CoA and 5 mM $MgCl_2$. The reaction was buffered with 100 mM NaH_2PO_4. PnG concentrations used in the assay varied and were adjusted to the ACPs present in the assay solution. PnG relative concentrations were 0.1, 0.25, 0.5, 1 and 2, whereas a PnG concentration of 1 equals the molar concentration of ACPs in the whole assembly line and 0.1 a tenth of the ACPs present. When bypassing PnA_{V4}, (2*Z*)-cyclohexanepropenyl-SNAC[4] was used as the diketide substrate analog. Immediately after mixing the reaction, 10 µL were added to 10 µL Methanol, as the negative control. The assay was run at 25°C and samples were taken after 2 h and overnight and quenched 1:1 with methanol.

Polyketide production assay with pretreated Pn PKS enzymes. The phoslactomycin production assays to assure the effect on production originates from continuous PnG activity was done as followed. It had to be assured, that the Pn PKS proteins are free of any mis-loaded acyl-residues bound to the ACPs. For this, PnA_{V4}, PnB and $PnC\text{-}TE_{DEBS}$ were expressed as previously described. After the Ni-NTA purification the proteins (including PnG) were desalted using PD-10 column and concentrated to a volume of approximately 1 mL. Each Pn PKS enzyme was treated separately with 5 µM Npt, 1 mM CoA, 5 µM PnG for 45 min in 10 mM $MgCl_2$, 100 mM NaH_2PO_4 at 28°C (assures complete activation of the ACPs to form *holo*-ACPs and complete removal of eventual acyl-residues bound to the ACPs). After this treatment the Pn PKS enzymes were cleared of PnG and CoA by size exclusion chromatography on a HiLoad 16/600 Superdex 200 pg column (300 mM NaCl, 25 mM NaH_2PO_4, pH 7.5). The assay was run at a volume of 50 µL and contained 5 µM Pn PKS proteins, 1 mM cyclohexanecarboxyl-CoA, 1 mM malonyl-CoA and 0.5 mM of each *α*-substituted malonyl-CoA derivatives, 5 mM NADPH and 100 mM NaH_2PO_4. PnG concentrations used in the assay varied and were adjusted to the number of ACPs present in the assay solution. Used PnG relative concentrations were 0.1, 0.25, 0.5, 1 and 2, whereas a PnG concentration of 1 equals the molar concentration of ACPs in the whole assembly line. Immediately after mixing the reaction, 10 µL were added to 10 µL Methanol, as the negative control. The assay was run at 25°C and samples were taken after 2 h and overnight and quenched 1:1 with methanol.

High resolution mass spectrometric analysis of the polyketide production assay. All samples were measured immediately or stored at -80°C. UPLC-high resolution MS analysis was carried out using an Agilent 6550 iFunnel QTOF LC-MS system equipped with an electrospray ionization source set to positive ionization mode. The analyte was separated on a RP-18 column (50 mm x 2.1 mm, particle size 1.7 µm, Kinetex EVO C18, Phenomenex) using a mobile phase system comprised 0.1% formic acid in water (Solvent A) and acetonitrile (Solvent B). Chromatographic separation was carried out using the following gradient condition at a flow rate of 250 µL min^{-1}: 0 min 5% B; 1 min 5% B, 6 min 95% B; 6.5 min 95% B; 7 min 5% B. The column oven was set to 40°C and auto sampler was maintained at 8 °C. Standard injection volume was 10 µL. Capillary voltage was set

at 3.5 kV and nitrogen gas was used as nebulizing (20 psig), drying (13 L min^{-1}, 225 °C) and sheath gas (12 L min^{-1}, 40°C). MS data were acquired with a scan range of 50-1200 m/z. LC-MS data were analyzed using MassHunter Qualitative Analysis software (Agilent).

Tandem mass spectrometry analysis of pentaketides. Determination of polyketides was performed using a HRES-LC-MS. The chromatographic separation was performed on an Thermo Scientific Vanquish HPLC System using a Kinetex Evo C18 column (50 x .12mm, 100 A, 1.7 µm, Phenomenex) equipped with a 20 x 2.1 mm guard column of similar specificity at a constant eluent flow rate of 0.25 ml/min and a column temperature of 40 °C with eluent A being 0.1% formic acid in water and eluent B being 0.1% formic acid in acetonitrile (Honeywell). The injection volume was 2 µl. The elution profile consisted of the following steps and linear gradients: 0 – 2 min constant at 0% B; 2 – 11 min from 0 to 100% B; 11 – 12 min constant at 100% B; 12 – 13 min from 100 to 0% B; 13 – 14 min constant at 0% B. A Thermo Scientific ID-X Orbitrap mass spectrometer was used in positive mode with an electrospray ionization source and the following conditions: ESI spray voltage 5000 V, sheath gas at 45 arbitrary units, auxiliary gas at 9 arbitrary units, sweep gas at 7 arbitrary units, ion transfer tube temperature at 300°C and vaporizer temperature at 325 °C. Scheduled targeted collision induced dissociation was performed on the five suspect molecules, applying a precursor ion scan at a mass range between 200 and 400 *m/z* with a mass resolution of 120000 using the orbitrap mass analyzer after quadrupole pre-isolation. Data dependent detection of MS2 spectra was performed at a normalized collision energy of 30% and an activation Time of 10 ms with an automatic definition of the scan range and a mass resolution (MS2) of 120000 using the orbitrap mass analyzer.

Kinetic analysis of PnG. ACP loading reactions contained 2 mM ACP, 10 µM Npt, 5 mM $MgCl_2$ and 2.4 mM acyl-CoA (exception butyl-CoA where 1.5 mM ACP and 1.6 mM butyl-CoA were used). After incubation of the loading reaction for 2 h at 28 °C the reaction was stored on ice. For the kinetic characterization of PnG mixtures with individual

concentrations of acyl-ACP (end concentrations between 1.8 mM and 5 µM) were treated with PnG (end concentration 20 nm in case of decarboxylated alkyl-ACPs and 500 nm in case of carboxylated acyl-ACPs) for individual durations between 1 and 20 minutes in a total volume of 10 µL. Subsequently the reactions were quenched with 10 µL of 20% (v/v) formic acid. The quenched reaction solutions were diluted with water to a final ACP concentration of 10 µM. 2 µL of the buffered protein solutions were desalted online using a Waters ACQUITY H-Class HPLC-system equipped with a MassPrep column (Waters). Desalted proteins were eluted into the ESI source of a Synapt G2Si mass spectrometer (Waters) by the following gradient of buffer A (water with 0.05% formic acid) and buffer B (acetonitrile with 0.045% formic acid) at a column temperature of 60 °C and a flow rate of 0.1 mL/min: Isocratic elution with 5% A for two minutes, followed by a linear gradient to 95% B within 8 minutes and holding 95% B for additional 4 minutes. Positive ions within the mass range of 500-5000 *m/z* were detected. Glu-Fibrinopeptide B was measured every 45 s for automatic mass drift correction. Averaged spectra were deconvoluted after baseline subtraction and eventually smoothing using MassLynx instrument software with MaxEnt1 extension.

PnG hydrolysis of CoA-thioesters. To assure that PnG does not hydrolyze polyketide substrates an Ultra High Performance Liquid Chromatography supported assay was done. For this, 5 µM PnG was incubated with 1 mM malonyl-, methylmalonyl-, ethylmalonyl-, butylmalonyl-, 3-methylbutylmalonyl-, hexylmalonyl- or cyclohexanecarboxyl-CoA in 50 mM NaH_2PO_4, pH 7.5 at 25°C. The control to test for background hydrolysis of the CoA esters did not contain PnG. Samples were taken at 0 min, 60 min and 240 min and quenched with formic acid (final in sample 10% v/v). The samples containing malonyl-, methylmalonyl- and ethylmalonyl-CoA were separated on Eurospher II 100-2 C18 column (100 x 2 mm, Knauer). Separation of these CoA-thioesters in the reaction samples was done with a gradient of 1.5 – 10% (v/v) acetonitrile in 10 mM potassium phosphate buffer (pH 6.8) over 6.5 min at a flow rate of 0.2 mL min^{-1} at 50 °C. The samples containing cyclohexanecarboxyl-, butylmalonyl-, 3-methylbutylmalonyl- and hexylmalonyl-CoA were separated on Luna 100-2 C18 column (150 x 4.6 mm, Phenomenex). Separation of these CoA-thioesters in the reaction

samples was done with a gradient of 2 – 18% (v/v) acetonitrile in 25 mM ammoniumformiate buffer (pH 8.1) over 9 min at a flow rate of 0.25 mL min^{-1} at 30 °C. To detect the CoA-thioesters via UV absorbance at 260 nm an InfinityLab Max-Light cartridge cell was used (60 mm detector length, Agilent Technologies Inc. Santa Clara, USA).

References

1. Fushimi, S.; Furihata, K.; Seto, H., Studies on new phosphate ester antifungal antibiotics phoslactomycins. II. Structure elucidation of phoslactomycins A to F. *The Journal of antibiotics* **1989,** *42* (7), 1026-36.
2. Chen, Y. L.; Zhao, J.; Liu, W.; Gao, J. F.; Tao, L. M.; Pan, H. X.; Tang, G. L., Identification of phoslactomycin biosynthetic gene clusters from Streptomyces platensis SAM-0654 and characterization of PnR1 and PnR2 as positive transcriptional regulators. *Gene* **2012,** *509* (2), 195-200.
3. Vogeli, B.; Geyer, K.; Gerlinger, P. D.; Benkstein, S.; Cortina, N. S.; Erb, T. J., Combining Promiscuous Acyl-CoA Oxidase and Enoyl-CoA Carboxylase/Reductases for Atypical Polyketide Extender Unit Biosynthesis. *Cell Chem Biol* **2018,** *25* (7), 833-839.e4.
4. Geyer, K.; Sundaram, S.; Susnik, P.; Koert, U.; Erb, T. J., Understanding Substrate Selectivity of Phoslactomycin Polyketide Synthase by Using Reconstituted in Vitro Systems. *Chembiochem : a European journal of chemical biology* **2020**.

Discussion

Polyketide synthases (PKS) are unique multifunctional enzymes with a modular organization. Each module incorporates a specific substrate and collaborates with downstream modules to assemble complex polyketides. Inspired by the modularity of PKS and the prospects of mixing and matching modules for the generation of *de novo* biosynthetic pathways, they have been targeted in synthetic biology approaches [1, 2]. Despite their promising potential for applying rational engineering, which would result in the production of new-to-nature chemicals, there has been a slowly progressing success. This is mainly due to an incomplete knowledge about their structure, the interaction between modules and the growing polyketide, in addition to a neglected in-depth biochemical characterization of whole polyketide synthases, modules, single domains and accessory enzymes. The difficulties in gaining full understanding of PKS, result in part from the little diversity of *in vitro* model systems. The investigation of multiple as well as more diverse model systems could aid in the better understanding of this class of enzymes. The labor-intensive synthesis of PKS substrates also impedes such in depth investigations. The characterization of wild type PKS, as well as engineered PKS is pivotal to the identification of bottlenecks caused by gatekeeping behaviors of single domains. Furthermore, particularly in the *in vivo* context, other enzymes might affect polyketide biosynthesis. This factor is often disregarded, when transferring *in vitro* knowledge to *in vivo* systems and leads to unsuccessful implementation of engineered PKS that seemed to be promising.

This work aims on tackling multiple of the here presented challenges. It demonstrates a resource efficient workflow for the chemo-enzymatic synthesis of a plethora of extender units. This enables in-depth biochemical characterization of PKS. The under-studied phoslactomycin (Pn) PKS was selected for *in vitro* reconstruction. It exhibits a unique substrate selectivity and complementing characteristics to preexisting model systems. Challenging the *in vitro* Pn PKS system with a versatile set of extender units led to the identification of a promiscuous module and the derivatization of Pn PKS derived polyketides. Furthermore, these assays allowed for the identification of bottlenecks in the naturally promiscuous wild type system, that cause an impeded incorporation of non-native extender units. Detailed kinetic analysis of the AT

(acyltransferase) domains in this system resulted in a deep understanding of extender unit selection. Finally, the TEII (type II thioesterase) an accessory enzyme from the phoslactomycin PKS, was identified to have a polyketide-production promoting effect, while simultaneously impacting polyketide diversification.

Workflow for the production of versatile extender units

The first challenge for thorough studies of PKS engineering, as well as for fundamental understanding of substrate selection and incorporation in polyketides, is the supply of extender units. An easily accessible and efficient supply path of diverse substrates with the correct stereochemistry is required. The common route for polyketide extender unit synthesis is ligation of malonic acid derivatives to CoA-SH by the ligase MatB [3]. While this is highly efficient with malonate and methylmalonate, this route is limited by the narrow substrate specificity of MatB for dicarboxylic acids (e.g. malonate) with short to no *α*-substitutions. Engineering of MatB resulted in broadening of its substrate spectrum from malonate- and methylmalonate towards for instance butylmalonate [3, 4]. After resolving this bottleneck, the ligation reaction remained to be limited by the availability of *α*-substituted malonic acid derivatives, as they need to be commercially accessible or laboriously synthesized by acylation of Meldrums acid [5, 6]. Finally, the reaction products of MatB are in the (2*R)* configuration, which is not accepted as substrate by PKSs. Using MatB synthesized extender units therefore additionally requires the supply of an epimerase that is capable of providing the (2*S*) configuration of the different extender units, which would allow polyketide biosynthesis.

In **chapter I** of this work, six ECR (enoyl-CoA carboxylase/reductase) homologues including one previously reported promiscuous, engineered ECR were investigated for their ability to produce a set of alkyl-malonyl-CoA derivatives with varying chain length, branching pattern and aromatic substituents. The benefits of ECR catalyzed extender unit synthesis are i) the availability to use a multitude of *α,β*-unsaturated acids that can chemically be coupled to CoA, ii) the native involvement and importance of ECR-II homologues in extender unit biosynthesis, iii) the immediate production of (2*S*) configured PKS substrates and iv) their high native substrate promiscuity [7]. With detailed kinetic analysis, $CcrC_{PAG}$, an engineered ECR-I from *Caulobacter crescentus* [7], was identified as

the homologue best-suited for polyketide extender unit biosynthesis, as it showed a pronounced substrate promiscuity and resulted in the highest carboxylation yields. Usage of this ECR circumvents the limitations of *α*-substitutions given in the case of MatB. Yet, the reductive side reaction of ECRs led to the accumulation of saturated byproducts, particularly phenylpropionyl-CoA, reducing overall product yield. To ensure efficient use of all resources, a system for byproduct recycling was sought. The previously described acyl-CoA oxidase Acx4 from *Arabidopsis thaliana* [8], which introduces the *α,β*-olefin group required for the ECR catalyzed carboxylation, was employed for biocatalytic proofreading. By recycling the saturated ECR reaction byproducts, the yield for benzylmalonyl-CoA production could be increased from 19% to 73% thereby reducing costs of this approach through efficient usage of CoA-SH.

Remarkably, introduction of the oxidase opened an additional path for extender unit biosynthesis, as Acx4 enabled the usage saturated alkyl-CoA thioesters as starting substrates. This allows additionally to employ *α,β*-unsaturated acids, for the exploitation of commercially available saturated carboxylic acids. After establishing a workflow for synthesis at preparative scale, these extender units were used in substrate competition assays with the DEBS *in vitro* system. Extender unit incorporation could be observed and previously reported AT mutations that enhance promiscuity were verified.

The synthesis of alternative extender units is however often hindered by the efficiency of the chemical coupling of CoA-SH to carboxylic acids, or acids carrying further functional groups, which can result in double activation with CoA or intramolecular cyclization. Enzymatic coupling could help to resolve this bottleneck, as for instance a double activation with CoA is highly unlikely through enzyme activation. An enzyme class that might be suited for conducting the ligation reaction are the middle chain fatty acyl-CoA ligases (FACL), which provide long chain acyl-CoA thioesters. These enzymes are naturally involved in the biosynthesis of some rare PKS extender units, where their activity is coupled to the ECR carboxylation reaction [9]. An additional point to add extender unit diversity is halogenation. Halogenation of natural products can lead to an increase in bioactivity, making such substrates of high interest for incorporation studies [10, 11]. The ECR from salinosporamide PKS, SalG, catalyzes the carboxylation of 4-chlorocrotonyl-CoA, resulting in chloroethylmalonyl-CoA [12]. The modularity of the in chapter I described

extender unit synthesis route allows for the combination of FACL ligase reaction to ECR-catalyzed carboxylation with, if required, previous Acx4-catalyzed oxidation. This can be further combined with FACL engineering towards a broader substrate tolerance. The incorporation of thus produced novel extender units may lead to immediate, site specific functionalization of polyketides resulting in the production of novel polyketides with potentially improved bioactivities.

In summary, the here presented ECR guided chemo-enzymatic synthesis enables the production a range of alkyl-malonyl-CoA extender units with the appropriate stereochemistry for polyketide biosynthesis. Additionally, biocatalytic proofreading allows for their production in high yields, paving the way for polyketide diversification and functional investigation of substrate selectivity.

Phoslactomycin PKS *in vitro* system

Retaining a high catalytic activity and efficiency in engineered PKS systems is required for efficient production of non-native polyketides. To accomplish this, a comprehensive biochemical understanding of the enzymatic mechanisms, for instance substrate specificity of domains and whole modules is pivotal. Studies showed that malonyl-CoA transferring AT domains are highly specific. On the other hand, alkyl-malonyl-CoA transferring AT domains have been shown to exhibit a broad substrate tolerance [13-17]. Most studies that focused on understanding and shifting substrate specificity, used naturally promiscuous AT domains as a starting point, which complicated the identification of specificity-conferring amino acid residues. The observation that multiple alkyl-malonyl-CoA transferring AT domains were capable of transacylating non-native substrates leads to the assumption that this is a general property of alkyl-malonyl-CoA transacylating AT domains.

Studying specificity, as well as promiscuity on the level of extender unit selection, requires a system, that ideally displays both behaviors. To tackle this problem, a new PKS model was sought, which exhibits both extender unit specificity and promiscuity. In the native context, Pn PKS incorporates the rare starter unit cyclohexanecarboxyl-CoA, five molecules of malonyl-CoA and two molecules of ethylmalonyl-CoA per full elongation cycle. It thus offers the possibility to study AT-substrate specificity for different extender

units, enabling insightful comparisons of domain reaction parameters in one closed system. **Chapter II** describes the reconstitution of the first six modules of Pn PKS *in vitro* with controlled product release at different sites, which resulted in the production of tetra-, penta- and hexaketides. Substrate incorporation assays showed that malonyl-CoA incorporating modules exclusively accept their native substrate, while the ethylmalonyl-incorporating fourth extending module PnC, incorporates all tested *α*-substituted extender units (methyl-, ethyl-, butyl-, 3-methylbutyl-, hexyl- and benzyl-malonyl-CoA). This supports the general promiscuity of alkyl-malonyl-CoA transacylating AT domains. PnC had a preference for medium chain length and branched residues, while malonyl-CoA was efficiently excluded from incorporation. Other promiscuous AT domains showed discrimination against longer and particularly, branched alkyl-residues [15, 17-19], rendering PnC_AT a domain with a remarkable broad substrate spectrum. Among all PKS domains, ATs show the greatest potential for engineering. The newly gained knowledge about PnC_AT can be used for diversification of phoslactomycins by supplying alternative extender units during *in vivo* production. Furthermore, it is conceivable to use PnC_AT for domain exchanges with PKS that are carefully chosen according to known restrictions of the systems, resulting in the diversification of different polyketides [20].

Pentaketide structures with the native and non-native alkyl-residues were proposed based on Pn PKS biosynthetic logic and TE_{DEBS} preference for lactone formation [21]. The proposed structures were supported by tandem mass spectrometric analysis. Noteworthy, pentaketides with C-2 substitutions longer than the two-carbon chain of an ethyl-residue, were not found in the reduced form, indicating that PnC_KR (ketoreductase) does not accept these non-native polyketides as a substrate. Subsequently, PnD can extend the native and non-native pentaketides to the respective hexaketides. It shows a preference for short residues at the C-2 position of the pentaketide, most likely conveyed by PnD_KS (ketosynthase). Either substrates with C-2 substitutions longer than ethyl-residues, or non-reduced pentaketides are excluded by the KS domain, since these were detected in lower amounts than anticipated. Depending on which scenario, future engineering should target either the KS or the KR-domain. Limitations caused by KS gatekeeping have been explored in *trans*-AT PKS [22, 23] and identified as well as elucidated in other *cis*-AT PKS [17, 24, 25]. One of the biggest challenges

in successful PKS engineering, will be the identification of mutations that affect gatekeeping behavior, while at the same time not affecting selectivity of domains and overall catalytic activity. Studies aiming to modify the structure of phoslactomycin derived polyketides will thus require a deeper examination of the bottlenecks reported in this study and the identification and implementation of promiscuity-promoting or substrate preference-shifting mutations.

The AT domains from Pn PKS were analyzed in detail with a steady state kinetic assay established for this work. This assay couples CoA-SH release to NADH consumption by utilizing the succinyl-CoA synthetase complex SucC/SucD from *Escherichia coli* with the pyruvate kinase/lactic dehydrogenase complex. It relies exclusively on purified proteins and allows for a continuous photometric, as well as fluorometric monitoring of enzyme reactions with nanomolar AT concentrations, under multiple turnover conditions. The reaction parameters support the strict substrate specificity of the malonyl-CoA transferring AT domains, previously observed in the here presented polyketide production assays. These AT domains, from PnA, PnB and PnD, discriminate against non-native substrates, most likely by the accessibility of the active site, as substrates carrying *α*-substitutions may not fit into the binding pocket. In contrast, PnC_AT, which native substrate is ethylmalonyl-CoA, can hydrolyze and transacylate all tested extender units. Intriguingly, it transfers the malonyl-residue onto its cognate ACP, although no incorporation into the pentaketide could be observed. This indicates the need for a more elaborate mechanism of influencing substrate specificities, than simply spatial restrictions. In the first half reaction of transacylation, an acyl-*O*-AT intermediate is formed. Here, a broad substrate tolerance of PnC_AT was observed, as hydrolysis and thus covalent binding to the active site for all tested substrates could be detected. This supports the theory that substrate discrimination does not take place during the acyl-*O*-AT intermediate formation [26]. In the second half reaction the acyl-residue is either hydrolytically released or transferred onto the ACP. The hydrolysis rate was virtually the same for all acyl-residues, implying that hydrolysis does not impact substrate selectivity. The transacylation reaction however, varied strongly depending on which acyl-residue is being transferred onto the ACP. For example, a hundred-fold higher transacylation rate was observed for the native substrate ethylmalonyl-CoA, over the non-native substrate

malonyl-CoA. The measured transacylation rates correlate with the incorporation patterns observed in substrate competition assays and show that specificity is controlled in the second half reaction by differential rates of transacylation.

Together, the results presented here, highlight Pn PKS as a powerful model to study PKS enzymology. The *in vitro* reconstitution of Pn PKS led to the identification of substrate preferences of single domains towards extender units, as well as growing polyketide chains. Limitations caused by the KR and KS domains towards non-native polyketides could be identified. With the ability to synthesize and supply a wide range of extender units PnC_AT could be identified as a highly promiscuous AT domain with the potential use for phoslactomycin derivatization and construction of chimeric PKS. In depth kinetic characterization of PnC_AT domain allowed decrypting of the mechanism behind extender unit selection.

Type II thioesterase PnG increases polyketide production efficiency

The *in vitro* reconstitution of Pn PKS not only enabled the characterization of this enzyme complex in detail but also the effects of accessory enzymes on the megasynthase and on product formation. The study of PKS associated TEIIs is often restricted to assessing their substrate spectrum with thioester analogs or ACP tethered substrates. Until now, their *in vitro* study with cognate natural product biosynthetic enzymes, was limited to type II PKS and NRPS/PKS hybrids [27-30]. Therefore, their impact on type I PKS in the well-defined *in vitro* settings remained unstudied. **Chapter III** aims on filling this knowledge gap, by the purification and biochemical characterization of PnG, the TEII from the Pn PKS and analyzing its effect on the reconstituted Pn PKS system. Investigation of PnG reaction parameters with a set of (alkyl)malonyl- and alkyl-residues presented on two different Pn ACPs revealed no preference for ACPs or alkyl-substituents. On the other hand, a higher specificity constant for malonyl- over ethylmalonyl-ACP and for the decarboxylated alkyl- over their cognate (alkyl)malonyl-ACP counterparts, could be measured. The reaction parameters suggest an editing role of PnG. The preferred substrates represent mis-loaded alkyl-residues and dead-end decarboxylation products. The activity of PnG integrates therefore well into the predominant function of TEIIs, which is editing by removal of ACP modifications that would otherwise stall assembly lines [31, 32].

Addition of PnG to Pn PKS assays led to an increase of tetra- and pentaketide production. To exclude that this effect stems solely from the clearing of alkyl-residues that originate from mis-loading by *holo*-ACP forming transferases, Pn PKS enzymes were cleared from eventual alkyl residues, prior to polyketide production assays. Subsequently it could be shown that the aberrant structures originate from mis-loading during protein expression in *E. coli* BAP1 which co-expresses a phosphopantetheinyl transferase. Importantly, they also arise during polyketide biosynthesis, supporting decarboxylation without subsequent condensation [27]. It can be expected that PnG performs the same function in the *in vivo* context of the native producer *Streptomyces platensis*, where the promiscuous 4'phosphopantetheinyl transferases Npt, draws the prosthetic group from a pool of CoA-thioesters and free CoA-SH. In particular acetyl-CoA is present in high concentrations compared to free CoA-SH [28], resulting in a high probability of transferring alkylated prosthetic groups and hence producing inactive PKS enzymes. Thus, the activity of a TEII is required and crucial to assure the functioning of the cost intensive PKS enzymes *in vivo*. Concomitantly, supply of PnG to polyketide production assays resulted in a drastic increase of *in vitro* polyketide formation.

Additionally, to increasing the amount of native and non-native polyketides, PnG caused a relative increase of the native product towards the non-native products in substrate competition assays, referred to as a product profile shaping. This effect is likely caused by an increased exposure time of non-native substrates on the surface of the ACP, as a result of missing interactions with downstream domains or loose interaction of the substrate with the ACP itself. Due to the increased exposure, PnG can access and remove the residues that block the assembly line with higher probability. An effect that becomes more pronounced with increasing PnG concentration. Engineering the incorporation of non-native substrates should focus on optimizing the interactions between substrate and domain, to increase PKS biosynthetic efficiency and to avoid removal of non-native substrates from the assembly line *in vitro* and *in vivo*. This is of particular importance as TEIIs are omnipresent in PKS and NRPS producing bacteria. PnG, and ultimately other editing TEIIs, could be used in similar assay setups like presented here, as an analytical tool to identify beneficial, interaction- and incorporation-promoting mutations. If a substrate is efficiently incorporated, addition of

PnG in increasing concentrations will result in a shaping of the product profile towards this (non-native) product.

The high substrate tolerance of PnG, as shown by kinetic analysis and polyketide production assays supports the low specificity model for TEIIs [29, 30]. To probe its promiscuity further, it was tested whether PnG can replace the terminal TE_{DEBS} on PnC. Indeed, similar levels of pentaketides were reached, indicating that PnG can release these polyketides and substitute TEIs, while simultaneously increasing the polyketide production. PnG could potentially be used in other systems for chain release to circumvent restrictions caused by i) substrate specificities of the terminal TE, ii) unsuccessful cloning of terminal TEs or iii) insolubility of chimeric proteins. Most TEIIs are responsible for the hydrolysis of acyl-thioesters, therefore releasing carboxylic acids. In the case of the phoslactomycin-derived pentaketides, the retention times of PnG- and TE_{DEBS}-released products are the same. Combined with the tandem mass spectrometric analysis, this indicates that these products contain the lactone ring formed most likely by spontaneous ring-closure after hydrolytic release.

Similar to the pentaketide assays described before, additional tests aimed on the ability of PnG to replace PnD tethered TE_{DEBS}. In the pentaketide assays, termination without TE_{DEBS} resulted in only approximately 5% of the product formation. This supports the importance of TE_{DEBS} for product release in that particular system and furthermore shows that substrates with C-2 substitutions with varying length can effectively be released by TE_{DEBS}. In contrast to the strong decrease of pentaketide formation in the absence of TE_{DEBS}, hexaketide production was two-fold increased. TE_{DEBS} has been described as a promiscuous thioesterase. It is however optimized for the macrocyclization of substrates with a C-2 methyl and C-3 hydroxyl group, in the presence of an appropriately placed distal hydroxyl group [21, 31, 32]. The efficient release of Pn PKS native and non-native pentaketides shows that the length of the C-2 substitution and (3*R*) configuration does not impair chain release. In the case of the phoslactomycin hexaketide, no C-2 substitution is present. A distal hydroxyl group for macrocyclization could however enable release by ring formation and therefore efficient hexaketide production. The impaired hexaketide production by PnD-TE_{DEBS} could be a result of protein mis-folding or slower TE_{DEBS} catalyzed product release due to the absence of a

C-2 substitution. It is further conceivable that TE_{DEBS} competes for the substrate with the dehydratase from PnD, which leads to a decreased assembly line efficiency.

Upon addition of PnG, production of the hexaketide was strongly increased, extending its polyketide releasing activity to the phoslactomycin derived hexaketide. Strikingly, addition of PnG had the strongest effect in the absence of TE_{DEBS}, supporting its detrimental effect on the phoslactomycin hexaketide system. The observation that the terminal TE can be a bottleneck for polyketide production was also made in the pikromycin system [33]. This highlights the importance of evaluating the activity and substrate spectrum of product-releasing terminal TEs in chimeric or engineered systems, as well as in mutasynthon approaches.

The work presented here showcases the production-promoting effect of PnG for increasing modular type I PKS *in vitro* polyketide biosynthesis. It was shown that PnG can be used for the release of final products, without a pronounced specificity towards structural features, with the ability of replacing terminal thioesterases. To use PnG for product release, it is important to consider that many polyketides require cyclization to result in bioactive compounds. It cannot be expected that PnG released products regio-selectively cyclize by spontaneous ring-formation. Thus, for ring-closure, a suited, terminal TE should be identified after the optimization of the PKS.

Studies focusing on the *in vivo* production of phoslactomycins in the native and heterologous hosts should include PnG activity, particularly when applying PKS and metabolic engineering. Disregarding the function of TEIIs could yield in unsuccessful implementation of mutations or strongly reduced product titers. Mediated by PnGs great substrate tolerance, it is a valuable tool to test polyketide production of truncated, chimeric and engineered PKS. PnG does not show the strong substrate specificity-bias of a terminal thioesterase and simultaneously increases polyketide production.

Future PKS research

Engineering of type I PKS by abstracting and subsequent recombining modules and domains, has the potential to produce a multitude of chemicals of interest. This includes compounds that are not normally produced biologically and only with great synthetic

chemistry efforts. Despite the advances made in the field of PKS engineering, it remains to be a long road until the full potential of PKS synthetic biology can be unleashed.

There is a strong need for the characterization of structural and conformational dynamics of PKS domains during biosynthesis. Particularly, regarding specific module-module interactions during the different steps of polyketide assembly and the interaction during polyketide-chain passage. A further focus will also be on the classical *in vitro* enzyme characterization, as presented in this dissertation, to broaden the collection on kinetic parameters and expand the enzyme toolbox. Carefully selected PKS, such as phoslactomycin PKS, can function as testbeds for detailed probing of the impact of mutations, domain swaps and mix-and-match approaches.

Understanding of the structural, mechanistic, biochemical but also evolutionary background is fundamental for to PKS pathway refactoring. Research on large PKS systems revealed evolutionary relationships between the domains within one system. In more detail, processing domains are more closely related to the downstream, than to the upstream KS and tend to co-evolve. This has led to a revision of the canonical module boundaries, resulting in a redefinition thereof [34]. Respecting these relationships and new boundaries for the construction of chimeric PKS has resulted in assembly lines retaining high catalytic efficiency [35]. Bioinformatic tools are having an increasing importance for the analysis and selection of modules suited for joining in chimeric PKS pathways. The generation of reliable computational tools will be based on wet-lab experimental and structural data. The constant feeding of newly gained data will be vital for automatized PKS refactoring [36]. If such automated PKS refactoring, that yields in the efficient production of desired compounds is achieved, further improvements can be gained by applying directed evolution combined with high throughput screening [20].

For many compounds, generation of the carbon backbone is sufficient to yield in the final product. Production of, for instance specialty chemicals by the generation of chimeric PKS was shown, that can replace their petroleum based chemical synthesis [37]. Extending this product palette could eventually lead to a reduction of anthropogenic emissions. For the generation of small molecule therapeutics, post-synthesis tailoring reactions are often necessary to create bioactivity. Identification, and modularization of

tailoring enzymes, will need to be addressed as a step following the production of fully functional, new-to-nature polyketide synthases.

References

1. Newman, D. J.; Cragg, G. M.; Snader, K. M., The influence of natural products upon drug discovery. *Natural product reports* **2000,** *17* (3), 215-234.
2. Newman, D. J.; Cragg, G. M., Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *Journal of Natural Products* **2020,** *83* (3), 770-803.
3. Koryakina, I.; Williams, G. J., Mutant malonyl-CoA synthetases with altered specificity for polyketide synthase extender unit generation. *Chembiochem : a European journal of chemical biology* **2011,** *12* (15), 2289-93.
4. Crosby, H. A.; Rank, K. C.; Rayment, I.; Escalante-Semerena, J. C., Structure-guided expansion of the substrate range of methylmalonyl coenzyme A synthetase (MatB) of Rhodopseudomonas palustris. *Applied and environmental microbiology* **2012,** *78* (18), 6619-6629.
5. Pohl, N. L.; Hans, M.; Lee, H. Y.; Kim, Y. S.; Cane, D. E.; Khosla, C., Remarkably broad substrate tolerance of malonyl-CoA synthetase, an enzyme capable of intracellular synthesis of polyketide precursors. *J Am Chem Soc* **2001,** *123* (24), 5822-3.
6. Bravo-Rodriguez, K.; Ismail-Ali, A. F.; Klopries, S.; Kushnir, S.; Ismail, S.; Fansa, E. K.; Wittinghofer, A.; Schulz, F.; Sanchez-Garcia, E., Predicted incorporation of non-native substrates by a polyketide synthase yields bioactive natural product derivatives. *Chembiochem : a European journal of chemical biology* **2014,** *15* (13), 1991-1997.
7. Peter, D. M.; Schada von Borzyskowski, L.; Kiefer, P.; Christen, P.; Vorholt, J. A.; Erb, T. J., Screening and engineering the synthetic potential of carboxylating reductases from central metabolism and polyketide biosynthesis. *Angewandte Chemie International Edition* **2015,** *54* (45), 13457-13461.
8. Schwander, T.; von Borzyskowski, L. S.; Burgener, S.; Cortina, N. S.; Erb, T. J., A synthetic pathway for the fixation of carbon dioxide in vitro. *Science* **2016,** *354* (6314), 900-904.
9. Miyazawa, T.; Takahashi, S.; Kawata, A.; Panthee, S.; Hayashi, T.; Shimizu, T.; Nogawa, T.; Osada, H., Identification of middle chain fatty acyl-CoA ligase responsible for the biosynthesis of 2-alkylmalonyl-CoAs for polyketide extender unit. *Journal of Biological Chemistry* **2015,** *290* (45), 26994-27011.
10. Eustáquio, A. S.; Moore, B. S., Mutasynthesis of fluorosalinosporamide, a potent and reversible inhibitor of the proteasome. *Angewandte Chemie* **2008,** *120* (21), 4000-4002.
11. Müller, K.; Faeh, C.; Diederich, F., Fluorine in pharmaceuticals: looking beyond intuition. *Science* **2007,** *317* (5846), 1881-1886.
12. Liu, Y.; Hazzard, C.; Eustáquio, A. S.; Reynolds, K. A.; Moore, B. S., Biosynthesis of salinosporamides from α, β-unsaturated fatty acids: implications for extending

polyketide synthase diversity. *Journal of the American Chemical Society* **2009,** *131* (30), 10376-10377.
13. Oliynyk, M.; Stark, C. B.; Bhatt, A.; Jones, M. A.; Hughes-Thomas, Z. A.; Wilkinson, C.; Oliynyk, Z.; Demydchuk, Y.; Staunton, J.; Leadlay, P. F., Analysis of the biosynthetic gene cluster for the polyether antibiotic monensin in Streptomyces cinnamonensis and evidence for the role of monB and monC genes in oxidative cyclization. *Molecular microbiology* **2003,** *49* (5), 1179-90.
14. Ismail-Ali, A.; Fansa, E. K.; Pryk, N.; Yahiaoui, S.; Kushnir, S.; Pflieger, M.; Wittinghofer, A.; Schulz, F., Biosynthesis-driven structure-activity relationship study of premonensin-derivatives. *Org Biomol Chem* **2016,** *14* (32), 7671-5.
15. Lowry, B.; Robbins, T.; Weng, C. H.; O'Brien, R. V.; Cane, D. E.; Khosla, C., In vitro reconstitution and analysis of the 6-deoxyerythronolide B synthase. *J Am Chem Soc* **2013,** *135* (45), 16809-12.
16. Bonnett, S. A.; Rath, C. M.; Shareef, A. R.; Joels, J. R.; Chemler, J. A.; Hakansson, K.; Reynolds, K.; Sherman, D. H., Acyl-CoA subunit selectivity in the pikromycin polyketide synthase PikAIV: steady-state kinetics and active-site occupancy analysis by FTICR-MS. *Chemistry & biology* **2011,** *18* (9), 1075-81.
17. Kalkreuter, E.; CroweTipton, J. M.; Lowell, A. N., Engineering the Substrate Specificity of a Modular Polyketide Synthase for Installation of Consecutive Non-Natural Extender Units. **2019,** *141* (5), 1961-1969.
18. Koryakina, I.; McArthur, J. B.; Draelos, M. M.; Williams, G. J., Promiscuity of a modular polyketide synthase towards natural and non-natural extender units. *Org Biomol Chem* **2013,** *11* (27), 4449-58.
19. Vogeli, B.; Geyer, K.; Gerlinger, P. D.; Benkstein, S.; Cortina, N. S.; Erb, T. J., Combining Promiscuous Acyl-CoA Oxidase and Enoyl-CoA Carboxylase/Reductases for Atypical Polyketide Extender Unit Biosynthesis. *Cell Chem Biol* **2018,** *25* (7), 833-839.e4.
20. Musiol-Kroll, E. M.; Wohlleben, W., Acyltransferases as tools for polyketide synthase engineering. *Antibiotics* **2018,** *7* (3), 62.
21. Weissman, K. J.; Smith, C. J.; Hanefeld, U.; Aggarwal, R.; Bycroft, M.; Staunton, J.; Leadlay, P. F., The Thioesterase of the Erythromycin-Producing Polyketide Synthase: Influence of Acyl Chain Structure on the Mode of Release of Substrate Analogues from the Acyl Enzyme Intermediates. *Angew Chem Int Ed Engl* **1998,** *37* (10), 1437-1440.
22. Nguyen, T.; Ishida, K.; Jenke-Kodama, H.; Dittmann, E.; Gurgui, C.; Hochmuth, T.; Taudien, S.; Platzer, M.; Hertweck, C.; Piel, J., Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection. *Nature biotechnology* **2008,** *26* (2), 225-233.
23. Jenner, M.; Frank, S.; Kampa, A.; Kohlhaas, C.; Pöplau, P.; Briggs, G. S.; Piel, J.; Oldham, N. J., Substrate specificity in ketosynthase domains from trans-AT polyketide synthases. *Angewandte Chemie International Edition* **2013,** *52* (4), 1143-1147.
24. Klaus, M.; Buyachuihan, L.; Grininger, M., Ketosynthase Domain Constrains the Design of Polyketide Synthases. *ACS chemical biology* **2020,** *15* (9), 2422-2432.
25. Murphy, A. C.; Hong, H.; Vance, S.; Broadhurst, R. W.; Leadlay, P. F., Broadening substrate specificity of a chain-extending ketosynthase through a single active-site mutation. *Chemical Communications* **2016,** *52* (54), 8373-8376.
26. Wang, Y.-Y.; Bai, L.-F.; Ran, X.-X.; Jiang, X.-H.; Wu, H.; Zhang, W.; Jin, M.-Y.; Li, Y.-Q.; Jiang, H., Biochemical characterization of a malonyl-specific acyltransferase

domain of FK506 biosynthetic polyketide synthase. *Protein and Peptide Letters* **2015,** *22* (1), 2-7.
27. Ad, O.; Thuronyi, B. W.; Chang, M. C., Elucidating the mechanism of fluorinated extender unit loading for improved production of fluorine-containing polyketides. *Proceedings of the National Academy of Sciences* **2017,** *114* (5), E660-E668.
28. Vallari, D. S.; Jackowski, S.; Rock, C. O., Regulation of pantothenate kinase by coenzyme A and its thioesters. *The Journal of biological chemistry* **1987,** *262* (6), 2468-71.
29. Heathcote, M. L.; Staunton, J.; Leadlay, P. F., Role of type II thioesterases: evidence for removal of short acyl chains produced by aberrant decarboxylation of chain extender units. *Chemistry & biology* **2001,** *8* (2), 207-20.
30. Claxton, H. B.; Akey, D. L.; Silver, M. K.; Admiraal, S. J.; Smith, J. L., Structure and functional analysis of RifR, the type II thioesterase from the rifamycin biosynthetic pathway. *The Journal of biological chemistry* **2009,** *284* (8), 5021-9.
31. Gokhale, R. S.; Hunziker, D.; Cane, D. E.; Khosla, C., Mechanism and specificity of the terminal thioesterase domain from the erythromycin polyketide synthase. *Chemistry & biology* **1999,** *6* (2), 117-125.
32. Aggarwal, R.; Caffrey, P.; Leadlay, P. F.; Smith, C. J.; Staunton, J., The thioesterase of the erythromycin-producing polyketide synthase: mechanistic studies in vitro to investigate its mode of action and substrate specificity. *Journal of the Chemical Society, Chemical Communications* **1995,** (15), 1519-1520.
33. Hansen, D. A.; Koch, A. A.; Sherman, D. H., Identification of a thioesterase bottleneck in the pikromycin pathway through full-module processing of unnatural pentaketides. *Journal of the American Chemical Society* **2017,** *139* (38), 13450-13455.
34. Keatinge-Clay, A. T., Polyketide synthase modules redefined. *Angewandte Chemie International Edition* **2017,** *56* (17), 4658-4660.
35. Nivina, A.; Yuet, K. P.; Hsu, J.; Khosla, C., Evolution and Diversity of Assembly-Line Polyketide Synthases: Focus Review. *Chemical reviews* **2019,** *119* (24), 12524-12547.
36. Eng, C. H.; Backman, T. W. H.; Bailey, C. B.; Magnan, C.; García Martín, H.; Katz, L.; Baldi, P.; Keasling, J. D., ClusterCAD: a computational platform for type I modular polyketide synthase design. *Nucleic Acids Res* **2018,** *46* (D1), D509-d515.
37. Hagen, A.; Poust, S.; Rond, T. d.; Fortman, J. L.; Katz, L.; Petzold, C. J.; Keasling, J. D., Engineering a polyketide synthase for in vitro production of adipic acid. *ACS synthetic biology* **2016,** *5* (1), 21-27.